GENOME-WIDE CENSUS AND EXPRESSION PROFILING OF CHICKEN NEUROPEPTIDE AND PROHORMONE CONVERTASE GENES

BY

KRISTIN RENEE DELFINO

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Animal Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2009

Urbana, Illinois

Adviser:

Associate Professor Sandra Rodriguez-Zas, Ph.D.

ABSTRACT

Neuropeptides regulate cell-cell signaling and influence many biological processes in vertebrates, including development, growth, and reproduction. The complex processing of neuropeptides from prohormone proteins by prohormone convertases, combined with the evolutionary distance between the chicken and mammalian species that have experienced extensive neuropeptide research, has led to the empirical confirmation of only 18 chicken prohormone proteins. To expand our knowledge of the neuropeptide and prohormone convertase gene complement, we performed an exhaustive survey of the chicken genomic, EST, and proteomic databases using a list of 95 neuropeptide and 7 prohormone convertase genes known in other species. Analysis of the EST resources and 22 microarray studies offered a comprehensive portrait of gene expression across multiple conditions. Five neuropeptide genes (apelin, cocaine- and amphetamine-regulated transcript protein, insulin-like 5, neuropeptide S, and neuropeptide B) previously unknown in chicken were identified and 62 genes were confirmed. Although most neuropeptide gene families known in the human are present in the chicken, several genes are not present in the chicken. Conversely, several chicken neuropeptide genes are absent from mammalian species, including C-RF amide, c-type natriuretic peptide 1 precursor, and renal natriuretic peptide. The prohormone convertases, with one exception, were found in the chicken genome. Bioinformatic models used to predict prohormone cleavages confirm that the processing of prohormone proteins into neuropeptides is similar between species. Neuropeptide genes are most frequently expressed in the brain and head, followed by the ovary and small intestine. Microarray analyses revealed that the expression of adrenomedullin, chromogranin-A, augurin, neuromedin-U, platelet-derived growth factor A and D, proenkephalin, relaxin-3, prepronociceptin, and insulin-like growth factor I was most susceptible

(*P-value* < 0.001) to changes in developmental stage, gender, and genetic line, among other conditions studied. Our complete survey and characterization facilitates understanding of neuropeptides genes in the chicken, an animal of importance to biomedical and agricultural research.

ACKNOWLEDGMENTS

This project would not have been possible without the support of many people. Many thanks to my adviser, Dr. Sandra Rodriguez-Zas, who read my numerous revisions and helped make some sense of the confusion. Also thanks to my committee members, Dr. Jonathan Sweedler and Dr. Roger Shanks,who offered guidance and support. And finally, thank you to my father and numerous friends who endured this long process with me, always offering support and love.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Many biological processes of health, well-being, and economic importance, including reproduction, development, growth, memory, and behavior are influenced by neuropeptides (Fricker, 2005; Hook et al., 2008). Neuropeptides are intercellular messengers that result from the post-translational cleavage and modification of prohormone proteins. Insights into these biological processes, molecular-assisted programs to select livestock species, and molecular-based tools to prevent, diagnose, and treat health disorders can be gained from the identification and functional characterization of neuropeptide genes and products. The first complete draft of the chicken genome was available in 2004 (International Chicken Genome Sequencing Consortium, 2004). However, few chicken neuropeptides are known, and few studies focus on profiling the expression of multiple chicken neuropeptide genes. In addition, no attempt has been made to predict chicken prohormone cleavages that could result in bioactive neuropeptides. The objective of this thesis research project was to address the previous limitations and conduct a genome-wide survey and functional annotation of chicken neuropeptide and associated enzyme genes. The aim of the following literature review is to present an overview of the various genome and bioinformatic tools and resources used to accomplish the objective.

CHAPTER 2: LITERATURE REVIEW

2.1 Chicken Genome

The chicken is the first agricultural animal to have its genome sequenced (International Chicken Genome Sequencing Consortium, 2004). The chicken genome is about one third the size of mammalian genomes, reflecting a substantial reduction in interspersed repeat content and segmental duplications (International Chicken Genome Sequencing Consortium, 2004). The chicken genome has a haploid content of 1.1×10^9 base pairs of DNA and is divided among 39 chromosomes, including 38 autosomes and one pair of sex chromosomes, with the female as the heterogametic sex (ZW female, ZZ male). Among the 39 chromosomes, 9 pairs of cytologically distinct macrochromosomes and 30 microchromosomes are present (Burt, 2007). Microchromosomes are a universal characteristic of all avian species. The typical avian karyotype contains usually around 30 pairs of small to tiny (between 23 and 7 Mb) microchromosomes. Recent work has shown that even though these chromosomes represent only 25% of the genome, they encode 50% of the genes (Burt, 2002).

In order to minimize heterozygosity and provide sequence for both the Z and W chromosomes, DNA of a single female of the inbred line of red jungle fowl (Gallus Gallus) was used by all sequencing libraries (International Chicken Genome Sequencing Consortium, 2004). A critical step in completing the sequencing of the chicken genome was high-throughput DNA sequencing of expressed sequence tags (EST) from tissue-specific cDNA libraries generated from several international projects (Cogburn et al., 2007). The assembly was generated from around 6.6 x coverage in whole genome shotgun reads, which is a combination of plasmid, fosmid, and

bacterial artificial chromosome (BAC)-end read pairs (International Chicken Genome Sequencing Consortium, 2004). In principle, BACs are used like plasmids; BACs are constructed from DNA of an organism (such as humans) and are inserted into a host bacterium. As the bacterium grows, BACs are replicated. Huge pieces of DNA can be easily replicated using BACs (Cogburn et al., 2007).

Chicken genomics has major applications and benefits in comparative genomics, evolutionary biology, development, and disease. The availability of new tools, such as whole genome gene expression arrays and single nucleotide polymorphism panels, coupled with the genome sequence, will enhance the use of the chicken as a model organism.

2.2 Chicken as a Model Organism

The chicken is considered to be the premier non-mammalian vertebrate model organism (McPherson et al., 2009). According to a review from Burt (2007), the roots of avian genomics go back to the emerging field of genetics over 100 years ago and many familiar terms, such as alleles, genetic linkage, and epistasis, were based upon the work with chicken morphological traits. Dodgson and Romanov (2004) stated that birds and mammals last had a common ancestor about 300 million years ago. From an evolutionary standpoint, the chicken can provide a valuable comparison to mammalian gene structure and function (Dodgson and Romanov, 2004); Basically, the assembly of the chicken genome sequence fills a gap in our knowledge in the evolution and conservation of vertebrate genomics (Burt, 2006). Extensive conservation of synteny between chickens and mammals was revealed through a comprehensive analysis of gene mapping data, along with analysis of sequences of vertebrate genomes (Burt, 2007). Chickens

have shown to retain an enormous amount of genetic diversity, and a large amount of these alleles have relevance to traits of interest in human biology (Dodgson and Romanov, 2004).

The chicken is extremely useful as a model species in the areas of developmental biology, virology, oncology, and immunology (Dodgson and Romanov, 2004). Because embryonic development occurs *in ovo* rather than *in utero*, one has ready access to the chicken embryo, which allows for direct manipulation and biochemical analysis (McPherson et al., 2009). In the area of virology, chickens were used as hosts in the discovery of retroviruses. The avian leukosis viruses (ALV) continue to be among the most intensely studied retroviral models (McPherson et al., 2009). Similarly, the first tumor virus (Rous sarcoma virus) and first oncogene were derived from chicken-based research (Dodgson and Romanov, 2004). Analysis of the chicken immune system provided the first indications of the distinctions between T and B cells; the B cell nomenclature derives from their origin in the chicken bursa of Fabricius (McPherson et al., 2009). Humans and chickens are infected by many common or related pathogens and have multiple similar resistance and susceptibly mechanisms (Dodgson and Romanov, 2004). The chicken is also an ideal model for genetic mapping and quantitative trait loci (QTL) analysis. The chicken reproduces rapidly, allowing several generations to be generated quickly, and has a large number of highly inbred lines available (McPherson et al., 2009). Overall, the chicken is an ideal model organism with which to compare mammalian genome structure and evolution.

The availability of the chicken genome and annotated set of genes provides new opportunities for whole genome based gene association and gene expression-based investigations. The chicken genome resource enabled two main research paths that were explored in this thesis project: the

identification of all the neuropeptide genes (and associated enzymes) in the chicken genome and the use of microarray gene expression experiments to profile the expression of thousands of neuropeptide genes (and enzymes) simultaneously. A literature review that provides background information on both paths is hereby presented.

2.3 Neuropeptides

Neuropeptides are biologically active peptides typically 3-40 amino acids in length. Neuropeptides act as peptide neurotransmitters and peptide hormones. More than 100 neuropeptide genes have been reported across species and taxa (Southey et al., 2006a; Tegge et al., 2008). Peptide neurotransmitters mediate neurotransmission, while peptide hormones mediate cell-cell communication. The same neuropeptide often serves important roles both as a neurotransmitter in the nervous system and as a peptide hormone in the peripheral endocrine system (Hook et al., 2008). Neuropeptides have many diverse biological functions that affect almost every brain system and neuronal network, as well as influence development and behavior. Some functions include regulation of reproduction; growth; water and salt metabolism; temperature control; food and water intake; and cardiovascular, gastrointestinal, and respiratory control (Strand, 1999). For example, recent studies observed with mice suggest that adrenomedullin (ADML) has a neuroprotective function. Possibly, it plays a role in maintaining homeostasis under normal and stress conditions. (Fernández et al., 2008). Studies with mice and humans have shown that pro-melanin-concentrating hormone (MCH) is associated with food intake, and possibly obesity (Hervieu, 2003). Prepronociceptin (PNOC) is widely distributed in the central nervous system, and studies with humans have shown that it plays a role in memory process (NCBI, 2009). Neurotensin (NEUT) has shown to involved in the maintenance of gut

structure and function and in the regulation of fat metabolism in the chicken (Esposito et al., 1997).

Neuropeptides are formed from large protein precursors, prohormones, which are progressively split by specific proteolytic enzymes (Hook et al., 2008; Strand, 1999). Prohormone precursors require proteolytic processing to liberate the active neuropeptide. Proteolytic processing occurs at dibasic and monobasic sites (arginine or lysine), as well as multibasic sites (Duckert et al., 2004; Hummon et al., 2006; Southey et al., 2006b; 2008ab; 2009). The precursor proteins may contain one or multiple copies of the active neuropeptide (Hummon et al., 2006; Strand, 1999).

Proteolytic processing is a key process required for the biosynthesis of numerous active neuropeptides from their inactive precursors. Biosynthesis begins with the translation of the respective mRNAs to generate the preprohormone precursors. Proteolytic processing occurs cotranslationally at the rough endoplasmic reticulum where the NH_2-terminal signal peptide of the preprohormone is cleaved by a signal peptidase. The resulting prohormone is routed through the Golgi apparatus. There, it is packaged into newly formed secretory vesicles together with processing proteases. Proteolytic processing occurs as the secretory vesicle matures so that the mature secretory vesicle contains a processed, biologically active neuropeptide that awaits cellular stimuli for regulated secretion (Hummon et al., 2006; Hook et al., 2008). Some neuropeptides undergo posttranslational modifications that may change the biological activities of the peptides. Posttranslational modifications that may alter the activities of neuropeptides include disulfide bond formation, glycosylation, COOH-terminal α-amidation, phosphorylaton, sulfation, and acetylation (Hook et al., 2008).

2.4 Prohormone Convertases

Prohormone convertases (PCs) are a family of evolutionary conserved dibasic and monobasic specific Ca^{2+} -dependent serine proteases of enzymes that clip off the active segments from a peptide at single, specific basic residues or pairs of basic residues (Strand, 1999). All prohormone convertases consist of a single peptide, a propeptide at the N-terminal end, a highly conserved catalytic segment, and a well conserved 150 amino acid downstream region known as a homo P-doman (Strand, 1999). The catalytic domain is the site where cleavage occurs, and this site has specificity for prohormones with multi-basic amino acid characteristics. This family of enzymes resemble a bacterial protease called subtilisin, with extensive homologies in their N-terminal portions (Strand, 1999). The seven known PC family members are: PC1, PC2, PC3 (or furin), PC4, PC5, PC6, and PC7 (Baea et al., 2008). The N-terminal pre-regions are signal peptides that direct the precursors which are post translationally modified to generate the mature molecule (Duckert et al., 2004). Prohormone convertases have been reported in several invertebrates, and have been demonstrated to possess highly conserved features across much of the animal kingdom, suggesting a similar mechanism across species for cleaving prohormones (Morash et al., 2009).

Most research on neuropeptides has been conducted on the human, the rat, and the mouse. A simple search in the scientific literature database PubMed (http://www.ncbi.nlm.nih.gov/pubmed) for neuropeptide and human, rat, mouse, or chicken resulted in 18,553, 16,312, 6,819, and 2,530 hits respectively. Although the number of neuropeptide literature references in chicken was high, only 17 chicken neuropeptides had been experimentally confirmed as of September 2009. Information on a large number prohormone

and PC sequences, including nucleotide and amino acid sequences, corresponding neuropeptides, and references across multiple species, is available in numerous public databases. This information was used to construct a master list of neuropeptide and PC genes that was in turn used to search for homologues genes in the chicken genome. The next sections summarize the bioinformatics resources and tools used in this thesis research project to detect all chicken neuropeptide and PC genes. Bioinformatics uses "informatics" techniques such as databases, algorithms, computer science, and statistics in order to understand and solve problems arising from biological data.

2.5 Bioinformatic Resources

UniProt. The UniProt (Universal Protein Resources, http://www.uniprot.org/) database is a comprehensive catalog of protein sequences and functional annotations and is a central resource for storing and interconnecting information from large and disparate sources (The UniProt Consortium, 2006). UniProt encompasses the UniProt Knowledgebase (UniProtKB) that was used in this study.

UniProtKB. The UniProtKB (http://www.uniprot.org/) database is the central foundation for the collection of functional information on proteins with accurate, consistent, and rich annotation. Core data mandatory for each UniProtKB entry includes amino acid sequence, protein name or description, taxonomic data, and citation information (NCBI, 2009). Other annotation information, such as widely accepted biological ontologies, classifications, and cross-references, are also often added. UniProtKB consists of two sections, UniProtKB/Swiss-Prot, which contains

manually reviewed annotated entries, and UniProtKB/TrEMBL, which contains computer-annotated entries that are awaiting full manual annotation and review (NCBI, 2009). UniProtKB/Swiss-Prot contains annotated information on protein or peptide function, catalytic activity, subcellular location, disease, structure, and post translational modifications. During the annotation process, different reports for a single protein are merged in order to have minimal redundancy and to improve sequence reliability. Cross-references are provided to various underlying nucleotide sequence sources and other useful databases, such as organism-specific, domain, family, and disease databases (The UniProt Consortium, 2006). UniProtKB/TrEMBL contains high quality computationally analyzed records that are enhanced by automatic annotation and classification. This database contains the translations of all coding sequences present in European Molecular Biology Laboratory (EMBL), GenBank of NCBI, and DNA Data Bank of Japan (DDBJ) Nucleotide Sequence Databases. According to defined annotation priorities, records are selected for full manual annotation and integration into UniProtKB/Swiss-Prot. Once in UniProtKB/Swiss-Prot, a protein entry is removed from UniProtKB/TrEMBL (The UniProt Consortium, 2006). Information from both Swiss-Prot and TrEMBL databases in UniProt (release 15.8, September 22, 2009) was considered in this study.

GenBank. GenBank (http://www.ncbi.nlm.nih.gov/Genbank/ ; release 173.0, August 15, 2009) is an annotated sequence database collection produced at the National Center for Biotechnology Information, or NCBI. GenBank consists of all publically available nucleotide sequences and their protein translations and is built by direct submissions from individual laboratories, as well as from bulk submissions from large-scale sequencing centers (NCBI, 2009). Bulk submissions include expressed sequence tags (ESTs), genome survey sequence (GSS), and other high-

throughput data (Benson et al., 2009). As of early 2009, GenBank includes over 95 billion nucleotide bases from more than 92 million individual sequences, with 16 million new sequences added in the past year. The contributions of the Whole Genome Shotgun (WGS) projects have supplemented this data to bring the total to 213 billion bases (Benson et al., 2009).

Sequences are classified using a comprehensive sequence-based taxonomy developed by NCBI and collaborators. Each GenBank entry consists of a description of the sequence, the scientific name and taxonomy of the source organism, bibliographic references, and a table of features listing areas of biological significance, such as coding regions and their protein translations, transcription units, repeat regions, and sits of mutations or modifications. To support specific sequencing strategies, files in GenBank are partitioned into 20 divisions, including high-throughput genomic (HTG) and EST (Benson et al., 2009).

High Throughput Genomic. The HTG division consists of transition unfinished large-scale genomic records, including sequences which are of draft quality but may contain 5'-UTRs and 3'-UTRs, partial-coding regions and introns (Benson et al., 2009). Records are characterized according to quality of data as Phase 0-3, with Phase 3 being a finished state. Once finished, HTG records are moved into GenBank (Benson et al., 2009). Quality assurance checks of each GenBank record include both a sequence and annotations. After quality control, the database record is assigned a unique identifier known as an Accession number. If part of a nucleotide sequence encodes a protein, a conceptual translation is annotated. A protein accession number is assigned to the translation product and is linked to a record for the protein sequence in NCBI's protein databases (NCBI, 2009).

EST. An expressed sequence tag, or EST, is a short sub-sequence of a transcribed cDNA sequence (Adams et al., 1991). Expressed sequence tags have been critical in gene discovery and sequence determination, as well as in identifying gene transcripts (Adams et al., 1991). This influence stems from the fact that ESTs are produced via one-shot sequencing of a cloned mRNA. Basically, several hundred base pairs are sequenced from a cDNA clone taken from a cDNA library. Due to the fact these clones consist of DNA complimentary to mRNA, ESTs represent portions of expressed genes (Adams et al., 1991).

Expressed sequence tags are a tool in which to refine predicted transcripts for genes, leading to prediction of their protein products and eventually of their function. For example, the situation in which ESTs are obtained, such as tissue or organ, gives information on the conditions in which the corresponding gene is acting. The design of precise probes for DNA microarray can be compiled from EST information that can then be used to determine gene expression (Adams et al., 1991). ESTs comprise over 30 billion nucleotide bases in GenBank (Benson et al., 2009). Through sequence homology (e.g. BLAST) searches, NCBI identifies all homologies for EST sequences and incorporates them into a dbEST database (http://www.ncbi.nlm.nih.gov/dbEST/) to be further incorporated into the UniGene database (Benson et al., 2009).

Genome. This database represents the current public build of the genome. The sequences in this database will have RefSeq accession numbers or type. These represent either contigs (from a clone based assembly) or supercontigs (from a whole genome shotgun or composite assembly).

The contigs in this database are from only the reference assembly. This database is generated at the time of a genome release (NCBI, 2009).

In this study HTGS, EST, and Genome databases stemmed from the chicken genome build 2.1.

Entrez Gene. This resource, also known as Gene, is NCBI's searchable gene-specific database comprised from RefSeq genomes. RefSeq is NCBI's reference sequence project aimed at collecting high quality, well annotated sequences of many types, including complete genomes, complete chromosomes, genomic regions, mRNAs, genome contigs, and proteins (Strand, 1999). The entries in this database result from a combination of both manual curation and automated analyses (Maglott et al., 2005). Each integer that is species specific is given a GeneID identifier (Maglott et al., 2005). Entrez Gene contains data for more than 3.2 million genes from over 4,500 organisms (Sayers et al., 2009). Focus is put on genomes that have been completely sequenced, have an active community to contribute gene specific information, or that are scheduled for intense sequence analysis (NCBI, 2009). Unique integer identifiers for genes and other loci for a subset of model organisms are provided in the database, as well as all information associated with those identifiers, thus, establishing a gene-to-sequence relationship. Maintained information includes nomenclature, chromosomal localization, gene products and their attributes, associated markers, phenotypes, and interactions. If available, results of the analyses that have been done on the sequence are provided. These results can include graphic summary of the genomic context; intro/exon structure and flanking genes; a link to a graphic view of the mRNA sequence; links to gene ontology and phenotypic information; links to corresponding protein

sequence data and conserved domains; and links to related resources (NCBI, 2009). The Gene repository also has links to the UniGene database.

UniGene. The UniGene (http://www.ncbi.nlm.nih.gov/unigene; build #41) database encompasses transcript sequences that appear (based on sequence similarity) to belong to the same transcription locus in the genome together with information on protein similarities, gene expression, cDNA clone reagents, and genomic location (NCBI, 2009). GenBank sequences are partitioned into clusters, each of which represents a unique gene. These clusters contain both mRNA sequences and ESTs, allowing representation of both known genes and putative genes based on expressed material that has been sequenced. To build clusters, all mRNA and EST sequences are compared, and overlapping sequences are assigned to the same cluster. Clusters containing full-length mRNA should have all ESTs deriving from the gene align with the mRNAs. In the case of clusters containing only ESTs, algorithms by which UniGene is built assemble the clusters out of overlapping ESTs in order to produce a picture of the gene from which it was putatively derived (Strand, 1999). Each successful cluster results in a UniGene entry containing information about the cluster including species, gene name, chromosome location, tissue distribution, and sequence information (NCBI, 2009). The information about the distribution of mRNA and ESTs expression available in the UniGene database allows to characterize the expression of genes across tissues, body parts, and developmental stages. Thus, UniGene is an excellent resource for designing tissue-specific microarrays for focused research projects. In our research, UniGene was used to profile the expression of chicken neuropeptide ESTs across tissues and stages. However, the gene expression profiles in UniGene are limited to variation across tissues and developmental stages. One of the objectives of the present study was

to compile a comprehensive annotation of the neuropeptide and PC gene expression. This annotation would span tissues and stages and also encompass other factors that can affect gene expression such as gender, genetic line, and various treatments. To accomplish this, a total of 22 chicken microarray experiments available in the Gene Expression Omnibus database (GEO) were analyzed and summarized.

Gene Expression Omnibus. The GEO database is a data repository and retrieval system for microarray and other forms of high-throughput molecular abundance data generated by the scientific community (Sayers et al., 2009; NCBI, 2009). A survey of the chicken microarray experiments available in GEO indicated that the Affymetrix in-situ synthesized oligonucleotide (GEO GPL3213) was the most commonly used chicken microarray platform. Furthermore, this platform had the highest representation of neuropeptide and PC genes including 53 neuropeptide and 5 PC genes transcripts. Therefore, all the chicken microarray experiments that used this platform available in GEO as of September 2009 were used in this study. The gene probes in the Affymetrix microarray platform are identified in the GEO database using the UniGene identifiers. The identification is based on the homology or similarity of the probe sequence to the gene sequence. The most commonly used program to align individual sequences (e.g. a microarray probe sequence) to a database of sequences (e.g. Gene or UniGene gene sequences) and identify homologues is BLAST.

2.6 Basic Local Alignment Tool

Basic Local Alignment Tool, or BLAST, is a set of similarity search programs designed to explore all of the available sequence databases. The program finds regions of similarity between

sequences and calculates a statistical significance score (NCBI, 2009). This score has a well-defined statistical interpretation, making real matches easier to distinguish from random background hits (Altschul et al., 1990). Nucleotide or protein sequences are compared to various sequence databases in order to understand functional and evolutionary relationships between sequences, as well as identify members of gene families (NCBI, 2009). The BLAST program uses a heuristic algorithm that seeks local alignment, rather than global alignment, allowing it to detect relationships among sequences that share only isolated regions of similarity (Altschul et al., 1990).

Determining common evolutionary ancestry, or finding homology between sequences, is fundamental to understanding the function of genomic sequences. Comparing two sequences for homology, one of which has known function, structure, or origin, allows inferences about the second unknown sequence to be made. Usually, homologous sequences share common elements of three-dimensional folding and secondary structure, although this is not always the case. When the statistical similarity between two sequences exceeds a threshold, homology is inferred. Similarity is measured through computing a local alignment, typically using a variant of the Smith-Waterman algorithm (a dynamic programming algorithm that evaluates all possible alignments between sequences to identify the best local alignment between two sequences) to find similar regions between two sequences. Point mutations or elementary changes that transform a sequence region in one sequence into a region in another are measured and then expressed as a probability that the score has arisen by chance. All possible alignments between two sequences are computed with a scoring scheme to compute an optimal local alignment.

Scoring of local alignment typically relies on a mutation data matrix and experimentally derived penalties for gaps (Cameron et al., 2004).

2.7 The Sequence Alignment E-Value

Statistical significance of each gapped alignment is expressed as an E-value. Three stages determine this value, an indicator of the degree of similarity between two aligned sequences.

1) A nominal score S is determined for each alignment using a mutation data scoring matrix and a function for penalizing gaps. The mutation scoring matrix is a matrix that places values to the alignment of identical (matches) and dissimilar (mismatches) nucleotides or amino acids. For example, amino acids that have different physical and chemical properties have negative or lower positive scores, meanwhile the alignment of identical amino acids have the highest scores. Also, matches between common amino acids have lower positive scores than matches of infrequent amino acids. Different mutation matrices have been created to address different evolutionary distances between sequences. For example, a mutation scoring matrix suitable for aligning evolutionary proximal sequences places more extreme positive and negative scores for the matches and mismatches, respectively; meanwhile a matrix suitable for distant sequences places less extreme scores for matches and mismatches because more changes (mutations) between the sequences are expected. Users can select from a range of mutation score matrices depending on the anticipated or desired distance between the query sequence (e.g. mouse neuropeptide Y gene sequence) and the sequences on the database (e.g. all the chicken sequences in the Genome database).

2) The nominal score is converted to a normalized S' score:

$$S' = (\lambda S - \ln K) \div \ln 2 \qquad [1]$$

where the values of λ and K are precomputed by random simulation for each scoring matrix and gap penalty combination. The purpose of the normalization is to adjust the nominal score by the scoring matrix used.

3) The normalized score in converted into an E-value:

$$E = Q \div 2^{S'} \qquad [2]$$

where Q is the search base size; Q is approximately equal to m×n, where m and n are the total number of residues in the query and the collection sequences, respectively (Cameron et al., 2004). The appeal of the E-value is that it is more comparable to a P-value than the normalized score. E-values are positive with low values (e.g. E-value < 1×10^{-5}) indicating strong evidence of sequence similarity. Furthermore, the E-value adjusts for the size of the database and query sequences, as longer sequences are expected to have higher scores and lower E-values by chance alone.

2.8 Gene Expression Microarray Technology

A DNA microarray is used to detect the presence and abundance of labeled nucleic acids in samples. Microarrays allow the simultaneous measurement of the expression of thousands of genes across treatments, developmental stages, or other conditions. The probes in the microarray can be oligonucleotides (or cDNAs) that are a perfect complementary match to a segment of the gene of interest. Gene expression describes the transcription of information contained within DNA into messenger RNA (mRNA) that is then translated into the proteins critical for the functions of cells (NCBI, 2009). By examining the amounts of mRNA that are produced by a cell, scientists are able to identify which genes are expressed, which in turn reveals how cells adapt to changes within and outside of the organism due to treatments or other conditions.

Changes in gene expression levels can be a suitable indicator of the changes in the abundance of protein (Stekel, 2003). Data from multiple microarray experiments are available in public databases like the Gene Expression Omnibus (GEO; http:// www.ncbi.nlm.nih.gov/geo), and the most commonly used chicken microarray platform is the Affymetrix in-situ synthesized oligonucleotide array.

Affymetrix In-Situ Synthesized Oligonucleotide Array Technology

Oligonucleotides are short (i.e. 20 to 80 nucleotides long) segments of RNA or DNA synthesized by cleaving longer segments, or by polymerizing individual nucleotide precursors. During in-situ synthesis, oligos are built up base-by-base on the surface of the array (Stekel, 2003). This occurs by a covalent reaction between the 5' hydroxyl group of the sugar of the last nucleotide to be attached and the phosphate group of the next nucleotide. A protective group on the 5' position of each added nucleotide prevents the addition of more than one base during each round of synthesis. Before moving on to the next round of synthesis, the protective group is converted to a hydroxyl group. Different methods can be used for this deprotection, with our interest focused on the Affymetrix technology.

Affymetrix GeneChip microarrays are high density oligonucleotide gene expression arrays widely used in genomics studies (Jiang et al., 2008). For deprotection, a photodeprotection method using masks is employed. Basically, masks allow light to pass to appropriate features, while keeping it from other features, in order to convert the protective group on the terminal nucleotide into a hydroxyl group where bases can then be added. Each step of synthesis requires a different mask. Photodeprotection has a coupling efficiency of about 95%, meaning about 95%

of nucleotides are successfully added at each step of synthesis. The longer the oligonucleotide being synthesized, the worse the yield of full length oligos. The composition of the final population of oligonucleotides is influenced by capping, which prevents further synthesis on a failed oligonucleotide resulting in all oligonucleotides on a feature to have the same start, but be of different lengths. On an Affymetrix microarray, each gene is represented by a probe set consisting of 11 different pairs of 25 base pair oligos covering features of the transcribed gene. A probe set consists of a series of probe pairs and represents an expressed transcript. Each pair consists of a perfect match (PM) and a mismatch (MM) oligonucleotide. The PM probe exactly matches the sequence of a particular standard genotype, while the MM differs in a single substitution in the central, 13th base and is designed to distinguish noise caused by non-specific hybridization from the specific hybridization signal. A single sample that has been previously labeled with a fluorescent dye (biotin) is hybridized to the microarray probes, the microarray is scanned, and an image of fluorescence intensities is obtained. Gene expression is determined by comparing the signal intensity from hybridization to probes complementary to the gene being measured with the signal intensity from hybridization to probes that contain mismatches; the signal from the mismatch probes are thought to represent cross-hybridization.

Affymetrix Microarray Experiment Protocol. There are four basic steps in using a microarray to measure gene expression in a sample: sample preparation and labeling; hybridization; washing; and image acquisition. The Affymetrix platform follows a specific protocol, which allows for easier comparison of results between laboratories than other platforms. First, RNA must be extracted from the tissue of interest. For labeling, one must construct a biotin-labeled complementary RNA for hybridizing to the GeneChip. The GeneChip anchors tens of thousands

of closely arrayed oligonucleotide probes on to the surface of solid substrates, where the process of recognition can be highly parallel (Xiao et al., 2005). In hybridization, DNA probes form heteroduplexes with labeled DNA (or RNA) via Watson-Crick base pairing. Slides are washed after hybridization to ensure only the target specifically bound to features on the array (DNA/RNA we are trying to measure) are left and to reduce cross-hybridization. Affymetrix has integrated its image-processing algorithms into the GeneChip experimental process. Overall, oligonucleotide arrays are powerful tools for monitoring gene expression and for resequencing genes.

Normalization. Microarray technology is susceptible to technical error even after following the recommended experimental protocols. Some of the systematic sources of variation can be removed through the process of normalization or preprocessing of the microarray gene expression intensities before analysis. Examples of these sources of technical error include arrays with higher overall expression levels than others with comparable samples due to scanning or labeling, and probes appearing more fluorescent than others due to the nucleotide composition. The most common normalization techniques applied to Affymetrix microarray experiments is the quantile normalization followed by the GC-Robust Multichip Average or GC-RMA (Cope et al., 2004; Speed et al., 1992). The quantile normalization results in all the arrays in an experiment having the same distribution, thus the same mean, variance, quantiles, and percentiles. This normalization does not remove the within gene changes in expression across treatments or samples but removes all systematic differences between arrays across genes. The RMA normalization provides a single gene expression value per gene and microarray by combining all the probe values within a gene. The CG-RMA adjusts the expression intensity by the

composition of bases and location of the bases on the probe. The rationale for this adjustment is that the hybridization bond between the complementary probe and sample bases depends on the labeling and base. The label typically binds to Cytosine C base, and this can interfere with the Cytosine-Guanine bond. However, the C- G bond is stronger than the Adenine-Thymine bond. An adjustment formula was developed to account for these factors in the normalization of the gene expression measurements. The normalizations of the microarray experiments in this study were implemented in Beehive (http://stagbeetle.animal.uiuc.edu/Beehive). Beehive is a public integrated suite of web-based tools to study gene expression data from microarray experiments (Smith et al., 2007).

2.9 Prediction of Prohormone Cleavage

Neuropeptides result from multiple cleavages of larger precursor proteins (Southey et al., 2006b). The complex and variable posttranslational processing of these precursor proteins makes identification of neuropeptides difficult. Cleavage precursors also tend to fluctuate across species, tissues, and developmental stages, as well as depend upon structural properties and the environment during peptide processing (Fricker, 2005; Tegge et al., 2008). Bioinformatics prediction of precursor cleavage sites can support effective biochemical characterization of neuropeptides (Tegge et al., 2008).

Comparative genomics and sequence homology are two common approaches for identifying precursor proteins (Southey et al., 2006a). Comparing sequences between well and poorly studied species can result in ample identification of precursor sequences, however, it may also result in inaccurate predictions of bioactive neuropeptides (Southey et al., 2006a). Sequence

homology may also result in inaccurate predictions because any one precursor processing step may depend on the presence of specific amino acids or combination of amino acids (or physiochemical properties) at specific sequence locations (Tegge et al., 2008). By considering multiple sequence locations, the identification of sequence motifs associated with cleavage can overcome some of the prior listed limitations (Southey et al., 2006b).

To gain a conclusive understanding of neuropeptide processing, there must be simultaneous consideration of multiple precursor families, species, and predictive approaches. Limited overlap between approaches and precursor data sets used to train and test cleavage prediction has hampered the characterization of precursor cleavages (Tegge et al., 2008). Currently, no phyla-specific models exist to predict cleavage of neuropeptides in the avian species.

The Logistic Regression model and Known Motif model are two suggested approaches to predict neuropeptide cleavage (Southey et al., 2008b). Analysis of these models results in predictions that are categorized into correct cleavage (true positive), incorrect cleavage (false positive), correct noncleavage (true negative), and incorrect noncleavage (false negative). The correct classification rate is the number of correctly predicted sites divided by the total number of sites. Sensitivity (the percentage of all cleaved sites that were correctly predicted) is the number of true positives divided by the total number of sites cleaved, and specificity (the percentage of all non-cleaved sites that were correctly predicted) is the number of true negatives divided by the total number of sites not cleaved (Tegge et al., 2008).

Logistic Regression model. The Logistic Regression model provides a probability of cleavage resulting from a linear combination of amino acids at different positions, relative to the cleavage site, weighted by their association with the cleavage (Southey et al., 2006b). This probability of cleavage, n_i, corresponding to the i^{th} window is described as

$$log[\pi_i(1-\pi_i)] = \sum_{j=1}^{p} \beta_j \, x_{ij}$$

where β_j is the regression coefficient associated with the j^{th} model term and x_{ij} denotes presence or absense of the j^{th} model term (j= 1 to p, amino acid or amino acid property at pericleavage locations) in the i^{th} window (Tegge et al., 2008). In this study, the human logistic regression model was used due to the human having a long-standing sequenced genome and extensive empirical validation of neuropeptide and precursor cleavage information on neuropeptides (Tegge et al., 2008). By using models trained on experimentally confirmed data in predicting precursor cleavages, some limitations of sequence homology and motif finding are overcome (Tegge et al., 2008).

Known Motif model. The Known Motif proposed by Southey et al., 2006b is comprised of several prevalent motifs associated with neuropeptide precursor cleavage, Xxx-Xxx-Lys-Lys↓, Xxx-Xxx-Lys-Arg ↓, Xxx-Xxx-Arg-Arg↓, Arg-Xxx-Xxx-Lys↓, and Arg-Xxx-Xxx-Arg↓, where ↓denotes cleavage and Xxx denotes any amino acid. Prohormone cleavages were predicted in this study using Neuropred.

Neuropred. Neuropred (http://neuroproteomics.scs.uiuc.edu/neuropred.html) is a web application used to predict the cleavage sites of neuropeptide precursors using logistic

regression models trained on experimentally verified cleavage information. Neuropred can also be used to calculate model accuracy indicators and neuropeptide masses from predicted cleavage cites (Southey et al., 2006a).

CHAPTER 3: GENOME-WIDE CENSUS EXPRESSION PROFILING OF CHICKEN NEUROPEPTIDE AND PROHORMONE CONVERTASE GENES[1]

3.1 INTRODUCTION

Neuropeptides encompass a wide range of small signalling peptides, such as neurotransmitters and peptide hormones, that regulate many biological processes, including reproduction, development, growth, memory, feeding, and behavior (Hook et al., 2008). These important intercellular messengers derive from larger prohormone proteins via a complex series of post-translational cleavages, spearheaded by prohormone convertases (PCs) and other post-translational modifications, which challenge their detection solely based on sequence homology to other more extensively studied species (Fricker, 2005; Hook et al., 2008). The chicken was the first avian genome sequenced (International Chicken Genome Consortium, 2004) and thus lacks closely related species with neuropeptide sequence information, although the song bird is currently being sequenced and annotated. The availability of the genome sequence allows one to uncover genes with limited or no empirical confirmation using bioinformatics tools, and the growing number of gene expression microarray experiments supports the functional annotation of these genes (Cogburn et al., 2003).

Although approximately 95 neuropeptide genes that code for prohormones have been identified in human, only 65 of these genes have been reported or predicted from the chicken genome and, in addition, prohormone peptide YY has only been reported at the protein level. The incomplete status of the chicken neuropeptide and PC gene complement is a notable deficiency considering the well-recognized status of the chicken as a model organism in biomedical and agricultural

[1]Parts of this chapter have been accepted for publication in Neuropeptides and the manuscript (Delfino, K., Southey, B., Sweedler, J., and Rodriguez-Zas, S. Genome-wide census expression profiling of chicken neuropeptide and prohormone convertase genes) is In Press as of November 20, 2009.

research (Stern, 2005). The role and expression patterns of a small percentage of these neuropeptides have been explored in chicken. Insulin-like growth factor 1 (**IGF1**) has a role in chicken fetal growth, as well as axonal growth and myelination (Duclos, 2005); Bennett et al. (2006) identified polymorphisms in IGF1 and insulin (**INS**) associated with weight at 5 weeks and 55 weeks in a layer-broiler cross in chickens; Zhou et al. (2005) reported significant associations between IGF1 and bone size and strength in 8-week old female and male chickens. Vasoactive intestinal peptide (**VIP**) relaxes the smooth muscle of trachea, stomach, and gall bladder. Jozsa et al. (2006) demonstrated that the brain levels of VIP and pituitary adenylate cyclase-activating polypeptide (**PACA**) change in chicken and rats after food deprivation and concluded that the 2 peptides are differentially involved in feeding.

The public Gene Expression Omnibus (GEO) database contains multiple chicken microarray gene expression platforms, including many with more than 10,000 probes (e.g. GPL1731, GPL1461, GPL1836, GPL2719, GPL2863, GPL3213, GPL4993, GPL5618, GPL6049). Although some of these platforms include neuropeptide and PC gene probes, the incomplete knowledge of the chicken neuropeptide and PC gene complement has challenged the profiling of these genes. In addition, the ability of mass spectrometry experiments to detect and characterize neuropeptides is aided by the availability of accurate prohormone gene identification and annotation (Li and Sweedler, 2008).

The objective of this study was to obtain the first genome-wide census and functional annotation of chicken neuropeptide and PC genes. First, an exhaustive master list of known neuropeptide and PC genes in the human and chicken was constructed. Second, the master list was searched

against various complementary chicken genome databases. Third, neuropeptide and PC gene expressions were profiled using a database of approximate expression patterns inferred from EST sources and a set of 22 chicken microarray experiments. Lastly, cleavage sites on the prohormone protein sequences were predicted and compared to known neuropeptide sequences and associated cleavages.

3.2 METHODS

Detection of Chicken Neuropeptide and Convertase Genes

A search for neuropeptide and PC genes across the chicken genome (1.1 Mb, including 30 microchromosomes and 9 macrochromosomes) was undertaken. A master list of candidate genes was generated based on public databases and a literature review (Amare et al., 2006; Southey et al., 2008b; Southey et al., 2009). The candidates were first searched for among the sequences already available in the GenBank (release 173.0, August 15, 2009) and UniProt databases (release 15.8, September 22, 2009). To uncover genes not previously reported or with different nomenclature from that of the master list, the human prohormone gene sequences were aligned against three resources stemming from the chicken genome build 2.1, the genome (**Genome)**, the expressed sequence tag (**EST**), and the high throughput genome sequence (**HTGS**) databases available in NCBI. Sequence searches were implemented using the chicken NCBI BLAST website (http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=9031) with default parameters (BLOSUM62 scoring matrix and maximum E-value of 10) and no filtering of low complexity regions. To augment the likelihood of identifying functionally conserved homologues, the protein sequence was used as a query. The matches were screened based on the

alignment E-value and distribution of the alignment identities, close matches, mismatches, and gaps along the sequence. The matches were also screened for alignments to related genes in the same neuropeptide family.

Characterization of the Neuropeptide and Convertase Gene Expression Profiles

The expression patterns of neuropeptide and PC genes were obtained from two resources. One resource was the UniGene database (build #41) which includes the expression of chicken neuropeptide ESTs across tissues and maturation stages. The other resource was the GEO database which encompasses gene expression experiments that used chicken microarray platforms and included probes for neuropeptide and PC genes. Affymetrix Chicken Genome Array GPL3213 (http://www.affymetrix.com/support/technical/byproduct.affx?product=chicken) was selected among the chicken microarray platforms because it had the highest number of relevant probes, including 53 neuropeptide and 5 PC gene transcripts. In addition, the GPL3213 platform had the highest number of gene expression experiments (22) of all chicken platforms. This unique abundance enabled a comprehensive analysis of all the experiments and the identification of neuropeptide and PC gene expression patterns across a wide range of conditions. The studies were grouped into 6 classes: retina, heart and breast muscle, brain and head, liver and duodenum, oocyte and gonad, and other tissues; the number of studies (and GEO series identification) within each class was 4 (GSE6543, GSE7176, GSE11439, and GSE15382), 4 (GSE6843, GSE8693, GSE9251, and GSE15413), 4 (GSE6844, GSE6868, GSE8693, GSE12268), 4 (GSE6856, GSE8483, GSE15413-liver, GSE15413-duodenum), 2 (GSE8693, GSE10231), and 5 (GSE8010, GSE8016, GSE8018, GSE8483, GSE9884), respectively. Experiments GSE8693 (Ellegren et al., 2007) and GSE15413 included comparisons of

conditions across multiple tissues, and the samples corresponding to each tissue were analyzed separately to facilitate the interpretation of results.

Pre-processing and normalization of the microarray data was done using the Affy R package (Irizarry et al., 2009) and included the $\log_2$ transformation of the intensities and GC-robust multichip average normalization of expression measurements. The expression measurements of all the probes in the platform were analyzed, and the statistical significance of the differential expression was adjusted for multiple testing across all probes using the false discovery rate approach (Benjamini and Hochberg, 1995). The microarray analyses were done using Beehive (http://stagbeetle.animal.uiuc.edu/Beehive).

Prediction of Cleavage Sites

Several models have been proposed to predict the cleavage of prohormone proteins coded by neuropeptide genes (Hummon et al., 2003; Southey et al., 2006b; Tegge et al., 2008; Southey et al. 2008a; Southey et al. 2009). However, no cleavage model has been trained on avian species. The accuracy to predict avian cleavage sites from the "known motif" model (Southey et al., 2006b) and the logistic regression model trained on human sequences (Tegge et al., 2008) was evaluated using the 24 chicken prohormone sequences that have peptide information (and in most cases with signal peptide information) available in UniProt. Both models are available at NeuroPred (http://neuroproteomics.scs.uiuc.edu/neuropred.html; Southey et al., 2006a).

3.3 RESULTS AND DISCUSSION

Chicken Neuropeptide Genes

A master list of 95 potential chicken neuropeptide genes and 7 PC genes were identified from the literature review and the Gene, UniGene, and UniProt databases. Table 1 summarizes the distribution of the genes on the master list across the 3 databases used to compile already known chicken genes (Gene, UniProt, UniGene) and the 3 databases used to uncover previously unreported chicken genes (Genome, HTGS, and EST databases). Additional information on the genes in the master list is available in supplementary material Table S1.

A total of 62 chicken neuropeptide genes were present in the Gene database and among them, 49 had the corresponding complete or partial prohormone sequence in UniProt. This count includes augurin or esophageal cancer-related gene 4 protein (**ECRG4**); although not currently present in the Gene database, a region in a genomic contig on chromosome 1 (ref|NW_001471545.1|Gga1_WGA43_2) had been assigned as being similar to ECRG4. Further, this count excludes the peptide YY-like (**PYY**-like) gene because although this peptide is reported in UniProt (P29203), no evidence for the corresponding gene sequence was found in any of the NCBI databases. The inability to confirm the UniProt PYY-like entry in other databases prompted us to remove this peptide from known chicken peptides. The proportion of prohormones coded by neuropeptide genes in UniProt that had empirical evidence at the protein level was 0.37 (18/49, not including PYY) and the remaining prohormones had evidence at the transcript level, or were based on sequence similarity or predictions (Table 1). Of the 62 neuropeptide genes in the Gene database, 59 had a corresponding record in UniGene. The

absence of UniGene entries for motilin (**MOTI**), oxytocin (**NEU1**), and orexigenic neuropeptide QRFP (**OX26**) reflects the lack of ESTs reported for these neuropeptide genes.

Of the 62 neuropeptide gene sequences in UniGene, 8 were not located in the chicken Genome, HTGS, or EST databases. Gastrin (**GAST**) and PACA were not located in Genome, HTGS, and EST databases, meanwhile C-type natriuretic peptide (**ANFC**), chromogranin (**CMGA**), prolactin-releasing peptide (**PRRP**), parathyroid hormone-related protein (**PTHR**), neuropeptide VF precursor (**RFRP**), and secretogranin-1 **(SCG1**) were not located in the HTGS database. UniProt does not have records for GAST, CMGA, and SCG1 that would support the corresponding UniGene records. A likely explanation for the neuropeptide genes not located in either of the three databases is that the Genome assembly and EST libraries may be incomplete at the locations of these genes.

Chicken Prohormone Convertase Genes

Of the 7 PC enzymes in the master list, only PC 4 (**PCSK4**) was not reported in the Gene database, and only 2 PC genes were reported in the UniProt database (Table 1). The 6 chicken PCs in the Gene database were confirmed in the chicken Genome, HTGS, and EST databases. Five of the PC genes in the Gene database had a corresponding record in UniGene. In addition to the UniGene partial record for PC 1 (**PCSK1**, Gga.9357), UniGene has a record (Gga.31439) predicted from the genome that is annotated to be similar to PC 1, although no corresponding Gene record for this UniGene record is available. The alignment between these two sequences has an E-value of 9×10^{-154} and 94% identity. Our survey complements the currently limited work on PC in the chicken. Ling et al. (2004) detected PC1 and PC2 mRNA in multiple chicken tissues including heart, lung, gizzard, pancreas, spleen, bursa of Fabricius, kidney, adipose tissue,

skeletal muscle, pituitary gland, cerebrum, mid-brain and cerebullum; and Richards and McMurtry (2008) reported that PC2 mRNA was mostly expressed in pancreas and proventriculus, whereas PC1 mRNA was more expressed in duodenum and brain of chicken.

No suitable match for PCSK4 in the chicken genome was found suggesting that PCSK4 may have evolved after the split between chickens and mammals, because PCSK4 is found in multiple mammalian species. The absence of evidence for PCSK4 is noteworthy because this convertase plays an essential role in the process of fertilization and is located in testicular germ cells in mice (Gyamera-Acheampong et al., 2006). The differences in the reproductive biology of mammals and chicken and the overlap of processing of some convertases (Baea et al., 2008) bar the conclusion that the absence of PCSK4 may hinder any particular prohormone cleavage or presence of a particular neuropeptide in chicken.

Previously Unreported Genes Detected in Chicken

The bioinformatics approach uncovered evidence for 5 neuropeptide genes that have not been previously reported. Confirmatory evidence in the Genome, HTGS, and EST databases that was further validated using complementary resources supports the discovery of evidence for the chicken homologues to the neuropeptide genes apelin (**APEL**), cocaine- and amphetamine-regulated transcript protein (**CART**), insulin-like 5 (**INSL5**), neuropeptide S (**NPS**), and neuropeptide B (**NPB**). The multi-step strategy varied across sequences and depended on the strength of the evidence, quality of the genome sequence, and availability of homologue sequences to confirm these findings. In the first step, 2 criteria were used to rule the finding of a previously unreported gene in the 3 databases: a low E-value of the sequence alignment

(encompassing a high percentage of identities and similarities with a minimum percentage of mismatches and gaps), and conservation of the region encompassing the gene product known in human were required. The additional confirmatory steps unique to each gene and a brief review of the implications of our findings in the understanding of chicken physiology, production, and health follows. Supplementary materials Table S2 presents a detailed description of sequence alignments for each of the 5 neuropeptide genes from the 3 databases, and underlined is the region corresponding to the functional neuropeptide known in humans.

APEL. The region of the chicken genome that matched the human APEL sequence (NP_059109) contained many gaps that prevented complete identification of chicken APEL from the genome. An EST (BU323997) that has a good match (*E-value* = 0.034) to the human sequence was identified, but the resulting BLAST alignment between the human sequence and BU323997 had 2 sections indicating a probable frame-shift in the EST sequence. The alignment of BU323997 against a region of the chicken genome on chromosome 4 indicates that an extra nucleotide 'G' in the EST not present in the genome sequence is present (Table S2a). After removing this extra 'G' and non-coding sequence from the chicken EST, a chicken APEL sequence was predicted. The trace archives (NCBI Trace-Other database) was searched using the putative nucleic APEL sequence to overcome limitations of the genome assembly. There are only 2 matches to this putative sequence which correspond to 2 exons as expected from the genomic structure of the human APEL gene. The detection of APEL in the chicken is significant because this neuropeptide has been implicated in a variety of roles in humans, including cardiac function; drinking behavior; regulation of adiposity, lipid and energy metabolism; and gastric cell proliferation (Higuchi et al., 2007).

CART. No clear and prolonged match to the human CART sequence in the chicken genome was found, suggesting that the CART gene region was not fully assembled in the chicken. However, following the strategy used for APEL, an EST (BM490862) with a good match (*E-value* = 7×10^{-32}) to the human CART sequence was identified. Considering that the human sequence matched the third reading frame of the EST and that the first 2 nucleotides of the EST were 'TG', we hypothesized that the EST sequence could be missing the initial 'A' nucleotide that would code for the first methionine amino acid in the CART protein. The start of the sequence was confirmed using the chicken trace archives (Trace-Others), and visualization of the matches showed that the middle of CART is missing in the genome sequences. The chicken EST mapped to a genomic contig yet to be located on the chicken genome (NW_001476554.1|GgaUn_WGA14361_2). The significance of the discovery of CART in the chicken is highlighted by its known role in reward, feeding, changes in body weight and fat mass, and stress (Asnicar et al., 2001; Kuhar et al., 2002).

INSL5. The search for human INSL5 in the chicken genome results in a match (*E-value* = 0.11) on chromosome 8 that was not attributable to members of the relaxin or insulin gene families. The NCBI annotation to this genomic region is "similar to WD repeat domain 78" (WDR78; Gene identifier LOC429114). The Map View feature in NCBI indicated that the chicken LOC424701 and TCTEX1D1 Tctex1 domain containing 1 gene are adjacent to this gene which is remarkable because human INSL5 is located on the negative strand between these 2 genes. Although the human INSL5 match was insufficient to identify a chicken INSL5 gene, a putative sequence was obtained using 3 fish sequences in the relaxin gene family; *Danio rerio* (Zebrafish,

B1AAQ6_DANRE), *Fugu rubripes* (Japanese pufferfish, B1AAR5_FUGRU), and *Tetraodon nigroviridis* (Green puffer, B1AAR6_TETNG). The *Danio rerio* sequence is shorter than the other 2 but the overlap is extensive. The alignment of the longer sequences to the chicken genome produced 2 very good alignments (*E-values* = 5×10^{-20} and 2×10^{-16}, respectively) that are 1210 bp apart (start of both chicken genome matches are 17631700 and 17630490, respectively). Both matching regions are annotated in the genome region as WDR78, further confirming that the inaccurate annotation of the chicken WDR78 gene prevented the annotation of the chicken INSL5 gene. The identification of INSL5 in chicken is notable because it has been postulated that INSL5 plays a role in gut contractility, remodeling and repair of the gastrointestinal tract, and neuroendocrine signaling (Conklin et al., 1999; Dun et al., 2006; Haugaard-Jönsson et al., 2009).

NPS. A match (*E-value* = 2×10^{-15}) to human NPS was identified on chicken chromosome 6 (Table S2b). The chicken genomic region that matched the human NPS sequence (± 5000 bp) was extracted, and the Wise2 version 2.1.20 software (http://www.ebi.ac.uk/Tools/Wise2/index.html; Birney et al., 2004) was used to predict the coded protein on the extracted region. The predicted chicken protein matched the complete human NPS sequence present in UniProt (Table S2b). The uncovering of NPS in the chicken genome is noteworthy because of its role in behavior (e.g. anxiolytic action, hyperlocomotion, wakefulness, altered sleep behavior, panic disorder), and intake (Castro et al., 2009; Pape et al., 2009).

NPB. The search for human NPB in the chicken genome matched a "hypothetical protein" (Gene database identifier 769277 LOC769277) on chromosome 18 (*E-value* = 2×10^{-6}). The genome region corresponding to actual bioactive NPB peptide was conserved in the alignment. Furthermore, the UniGene entry associated with this hypothetical protein clusters it with "anaphase promoting complex subunit 11" representatives from other species including human, mouse, and zebra fish. The parsing of the chicken nucleic genomic region of the human match using Wise2 uncovered one gene with 2 exons. Exon 1 includes the neuropeptide and is well conserved with the human NPB gene that also has 2 exons. Similar analyses using the known NPB sequences of the zebra fish and salmon offered results consistent with those obtained using the human sequence. The alignment of the NPB prohormone sequences predicted from the chicken Trace-Other archives includes gaps indicating that incomplete genome assembly prevented the annotation of the chicken NPB gene. Our discovery of NPB in chicken is significant because the NPB/NPW neuropeptide system regulates energy homeostasis, pain, and emotion, and NPB exerts strong synergistic anorectic effects in mice when co-administered with CRF (Hondo et al., 2008; Aikawa et al., 2008).

Neuropeptide Genes Not Located In Either Chicken or Mammals

Evidence was insufficient to locate 26 neuropeptide genes from the master list in the chicken genome: intermedin (**ADM2**), natriuretic peptide B (**ANFB**), calcitonin-related polypeptide beta (**CALCB)**, cortistatin (**CORT**), galanin-like peptide (**GALP**), hepcidin (**HEPC**), insulin-like 3 and 6 (**INSL3** and **INSL6,** respectively), metastasis-suppressor KiSS-1 (**KISS1**), neuromedin S (**NMS**), neuropeptide FF (**NPFF**), neuropeptide W (**NPW**), proprotein convertase subtilisin/kexin type 1 inhibitor (**PCSK1N**), proenkephalin B (**PDYN**), putative peptide YY-2

(**PYY2**), pro-relaxin 1 and 2 (**REL1** and **REL2**, respectively), regulated endocrine-specific protein 18 (**RES18**), spexin (**SPXN**), tachykinin 4 (**TAC4**), parathyroid hormone 2 (**TIP39**), tachykinin 3 (**TKNK**), torsin family 2 member A isoform prosalusin (**TOR2X**), urocortin (**UCN1**), and urocortin 2 (**UCN2**). Although Gene and UniGene have an entry for chicken torsin family 2 member A (XP_415507, Gga.5228), the chicken sequence corresponds to a human torsin alternative splicing isoform (**TOR2A**) that does not code for the neuropeptide salusin, and thus was considered not located in the chicken genome. In contrast, 3 neuropeptide genes in the master list (C-RF amide, c-type natriuretic peptide 1 precursor, and renal natriuretic peptide) were present in chicken but were not located in mammalian species.

A remarkable finding is that the vast majority of the human genes not located in the chicken genome have at least 1 neuropeptide gene in the same family present in the chicken genome (Table 1). For example UCN1 and UCN2 are absent from the chicken genome, while urocortin 3 (**UCN3**) is present in the chicken genome. The exceptions to the presence of at least 1 member of the neuropeptide family in the chicken genome are CORT, HEPC, KISS1, NPFF, NPW, RES18, and SPXN. Burt (2007) noted the low number of genes identified in the chicken genome relative to the human genome and hypothesized that more duplication events occurred in the human lineage of some genes and more losses of other genes occurred in the avian lineage. Our results reinforce this hypothesis that some neuropeptide genes have undergone substantially lower gene duplication in the chicken compared to human.

Neuropeptide Gene Expression Across Tissues and Developmental Stages

The distribution of the expression of most neuropeptide and PC genes available in UniGene was used to gain an initial understanding of the expression profiles across tissues and developmental stages. All PC and neuropeptide genes in UniGene, with the exceptions of c-type natriuretic peptide 1 (**CNP1**), CRF, GAST, progonadoliberin 1 (**GON1**), pancreatic polypeptide (**PAHO)**, parathyroid hormone (**PTHY**), prothyroliberin (**TRH**), thyroid stimulating hormone subunit beta (**TSHB**), and urotensin 2 (**UTS2**), had expression information. For PCSK1, the corresponding UniGene record (Gga.9357) did not have expression information, but another UniGene record (Gga.31439), annotated as "similar to PC1", was used as proxy because of the availability of expression information and similarity to the PCSK1 sequence. Table 2 provides a summary of the expression profile of 51 neuropeptide and 6 PC genes across the 5 tissues and 2 stages with most frequent neuropeptide gene expression out of 19 tissues or body parts and 4 development or maturation stages. Supplementary materials Table S3 presents the distribution of expression across all 19 tissues and 4 stages.

The tissue or body part with highest number of neuropeptide gene expression reports (expressed in absolute number and percentage) was the brain (33, 65%), followed by head (21, 41%), ovary (18, 35%), small intestine (16, 31%), and heart (13, 25%). A similar distribution was observed for the PCs, with the brain and small intestine being the body parts with highest frequency of gene expression. These results are consistent with the role of neuropeptides in physiology, health, and behavior (Hook et al., 2008). The developmental-maturation stage with the most reports of neuropeptide gene expression was adult (42, 82%), followed by embryo (39, 76%). The neuropeptide genes with the highest number of reports of expression across tissues or body

parts were platelet-derived growth factor alpha polypeptide (**PDGFA**, 11, 58%), SCG1 (11, 58%), and ECRG4 (9, 47%). These results are consistent with neuropeptide research across species. For example, PDGFA is expressed in the seminiferous epithelium and interstitial mesenchymal cells, and studies with mice show it may play a role in cell proliferation, migration in osteoblastic cells, and in production of Leydig cells (Yang et al., 2008). Likewise, the tyrosine-sulfated secretory protein SCG1, found in a wide variety of peptidergic endocrine cells in mice, may play a role in the early phase of neoplastic progression (Lukinius et al., 2003). Also, the expression of ECRG4 in multiple tissues, including the heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas, suggests a role in the modulation of salt and energy homeostasis, cardiovascular function, and cerebral spinal fluid composition (Mirabeau et al., 2007; Mori et al., 2007).

Although the distribution of expression reports can be influenced by the imbalanced distribution of EST libraries and experimental interest across tissues, developmental stages, and neuropeptide genes, all chicken tissues and developmental stages had at least 1 neuropeptide gene with a UniGene expression report. This confirms the importance of neuropeptides on all aspects of chicken physiology, growth, reproduction, and health.

Expression Profiling Based On 22 Microarray Experiments

Although the information in UniGene offers a broad picture of the expression of neuropeptide and PC genes across a wide range of tissues and stages, additional conditions can influence the expression profile. To fully investigate the variation in neuropeptide and PC gene expression across a wide range of conditions and augment the understanding of the impact of these genes on

reproduction, health, growth, and other traits of importance to biomedical research and agricultural production, the information from a large number of microarray gene expression experiments investigating numerous conditions was mined.

We present results from the first simultaneous analysis of 22 microarray experiments to characterize the expression of neuropeptide genes and PCs across a wide range of conditions. The in-situ synthesized microarray platform selected has the highest representation of chicken neuropeptide and PC genes available in GEO and is most widely used. A total of 73 probes representing 53 neuropeptide and 5 PC genes was available in the platform. The experiments were broadly grouped into retina, heart and breast muscle, brain and hypothalamus, liver and duodenum, gonad and oocyte, chicken-quail comparisons, and other conditions. To facilitate the interpretation of the results, a summary of the experiments and their features (e.g., tissues, treatments, age, gender, genetic line) is presented in Table 3. More detailed descriptions of the experiments are available in supplementary material Table S4 and in the GEO database. Due to the multiple probes analyzed, a minimum false discovery rate multiple-test adjusted *P-value* < 0.05 threshold (corresponds to an approximate unadjusted *P-value* < 0.005) and a minimum fold-change equal to 1.25 was used to identify differentially expressed genes. Table 4 summarizes the number of probes with differential expression across experimental group and probes corresponding to the same gene. Supplementary materials Table S5 presents the detailed distribution of the differential expression level of each probe and experiment. A summary of the main findings by tissue group are described below.

Retina. The chicken is a well-established model for the human eye and retinal development and degeneration (Burt, 2007). Two independent microarray experiments GSE6543 (McGlinn et al., 2007) and GSE11439 (Schippert et al., 2008) investigated the effect of myopia and the lens in the retina in chicken, respectively. The neurotensin (**NEUT)** gene was significantly over-expressed in the treated samples relative to the control samples in both studies. In addition, glucagon (**GLUC**) and NEUT were over-expressed in the treated samples relative to the control in the GSE11439 experiment, and VIP was under-expressed in the treated samples relative to the control in the GSE6543 experiment. These findings confirm the important role of neuropeptides in vision in the chicken (Schwippert et al., 1998; Chapman and Debski, 1995). Likewise, from experiment GSE7176 (Rizzolo et al., 2007) that studied the chicken retina across embryo developmental stages, we uncovered that the expression of gene adrenomedullin (**ADML**) was significantly lower in embryonic day 7, or E7 (*P-value* $< 1 \times 10^{-6}$ and 0.29 average fold change) relative to more advanced ages (E10, E14, and E18). Both the ADML peptide and its mRNA have been detected in embryonic mice in the outer neuroblastic layer of the retina (Montuenga et al., 1997) and in the human retinal pigment epithelial cells, suggesting an important physiological role for ADML in eye development (Udono et al., 2000). On the other hand, the variation in the expression of VEGFC across developmental stages was significant (*P-value* $< 3.0 \times 10^{-4}$) but higher at E7 relative to E10, E14, and E18 (2.24 average fold change). Neuropeptide VEGFC is expressed in the retinal astrocytes and promotes both endothelial cell proliferation and migration (Alon et al., 1995; Stone et al., 1995; Pierce et al., 1996; Provis et al., 1997). Likewise, the expression had a significant fluctuation across stages (*P-value* $< 1.0 \times 10^{-16}$) with higher expression at E7 relative to E10, E14, and E18 (3.44 average fold change). This profile is consistent with studies in mice that have shown that the PDGFA receptor, which is

activated mainly by PDGFA located in retinal neurons, is expressed at all stages of maturation and is important for retinal astrocyte proliferation and migration (Mudhar et al., 1993; Fruttiger et al., 1996). Over-expression of PDGFA in transgenic mice causes a significant increase in retinal astrocytes, resulting in proportional overgrowth of the retinal vasculature (Fruttiger et al., 1996).

Experiment GSE15382 (Kubo and Nakagawa, 2009) aimed to investigate the potential impact of c-hairy1, a gene that inhibits neuronal differentiation on gene expression. Our analysis uncovered that samples with this gene exhibited over-expression of neuropeptide Y (**NPY**) (*P-value* < 3.0×10^{-2}, 11.31 fold change) and to a lesser extent, secretogranin 2 (**SCG2**, *P-value* < 2.0×10^{-2}, 1.13 fold change) relative to control samples. In addition, the differential expression between c-hairy1 and control was similar to that between c-hairy1 and Delta and Wnt2b, 2 other genes expected to also inhibit neuronal differentiation. Thus, c-hairy1 had a strong association with the expression of 2 neuropeptide genes that is not observed with the other 2 potential inhibitor genes. These results are consistent with reports that NPY, along with its receptors, are present in the retina of both mammalian and non-mammalian species (D'Angelo and Brecha, 2004) and that this neuropeptide may modulate the development of retinal circuitry in rats (Bagnoli et al., 2003). Recent experiments with rats have shown that NPY produces a 2-fold increase in retinal neural cell proliferation and promotes the proliferation of committed neural immature cells (Álvaro et al., 2008).

Heart and Breast Muscle. Neuropeptides can contract muscles, and the action of a neuropeptide is of significance in the control of antagonistic contractions (Cho et al., 1996). Experiments

GSE6843 (Itoh et al., 2007) and GSE8693 investigated gene expression in embryonic heart tissue, and our analyses did not detect differential expression among the neuropeptide genes studied between females and males, suggesting that the role of the neuropeptides may be equally important in both genders. On the other hand, neuropeptide gene expression can exhibit significant variation across breast muscle at different development stages. For example, the analysis of experiment GSE15413 uncovered numerous neuropeptide genes that were differentially expressed in the breast muscle between 7-d-old and just hatched (0-d-old) chickens. Specifically, prepronociceptin (**PNOC**, 2 probes), PDGFA, platelet-derived growth factor beta polypeptide (**PDGFB**), platelet derived growth factor D (**PDGFD**), ADML, and ECRG4 were over expressed in 0-d-old relative to 7-d-old chickens. These findings are consistent with studies that report platelet-derived growth factors are important in avian embryonic development (Van Den Akker et al., 2005). Conversely, proenkephalin (**PENK**) and IGF1 were over-expressed in 7-d-old relative to 0-d-old chickens. No neuropeptide gene was differentially expressed between male and female embryo heart samples at 18-d-old in GSE8693 or between male and female embryos at late stages of development in GSE6843.

Experiment GSE9251 (Zheng et al., 2009) profiled the expression of genes in the breast muscle of broiler and layer genetic lines at different ages. Our analysis found that PNOC appears to have a quadratic expression pattern, regardless of genetic line, because it is under-expressed in both broiler and layer at young (1-d-old) and old (6-wk-old to 8-wk-old) ages relative to intermediate ages (2-wk-old to 4-wk-old). Both lines had similar levels of fluctuation across ages (approximate *P-value* < 0.0005 and approximate maximum fold change 2.22). This result is consistent with the role of nociceptin in stimulating locomotion (Florin et al., 1997). The PNOC

gene is highly conserved within the mouse, rat, and human, and studies have shown that it is broadly expressed in the nervous system, primarily in the brain and spinal cord (Mollereau et al., 1996). For PRRP, SCG2, ANFC, musclin (**OSTN**), neuromedin-U (**NMU**), TSHB, TRH, somatostatin (**SMS**), islet amyloid polypeptide **(IAAP**), gastric inhibitory polypeptide **(GIP)**, and tachykinin, precursor 1 (**TKN1**), the expression at 1-d-old was lower than at older ages in both genetic lines, and the fold change did not differ significantly between lines (maximum *P-value* = 8.4×10^{-3}). The ECRG4 gene was highly differentially expressed (*P-value* < 1.0×10^{-16}) and had the highest expression in 1-d-old broiler and layers with both lines showing similar fold changes. Gene ADML exhibited differential expression (*P-value* < 1.0×10^{-16}) with the level at 1-d-old being higher than at older ages in broilers; meanwhile for layers, ADML is under-expressed in 1-d-old chickens relative to older ages, and the level is significantly different between the lines in 1 d-old chickens. The difference in expression between genetic lines may be associated with the angiogenic role of ADML, albeit weaker in chicken than in human and mouse (Martínez et al., 2006). The expression of IGF1 varied significantly across ages and lines (*P-value* < 2.4×10^{-13}) with 2-wk-old chickens having the highest expression among layers, with expression not varying across ages within broilers, and not showing a clear pattern, significant trend, or deviation at a particular time point. The maximum difference between lines (2.15 fold change over-expression in layers relative to broilers) was observed at 2-wk of age. For the IGF2 probe set, the minimum expression for both genetic lines was in 1-d-old chickens (*P-value* < 3.9×10^{-3}) and the lines did not differ in the level of expression at that age. This result does not support the hypothesis postulated by Wang et al. (2005) that IGF2 can be a candidate gene influencing growth and carcass traits, although their work only used broilers. The differential expression (*P-value* < 8.4×10^{-5}) of the vascular endothelial growth factor C (**VEGFC**) gene

exhibited the maximum at 2-wk-old across lines. The differential expression (*P-value* < 1.0×10^{-16}) of vascular endothelial growth factor D (**VEGFD**) encompassed the minimum and maximum expression in 1-d-old and 2-wk-old broilers, respectively. The level of expression in broilers is significantly lower than layers at 1-d-old (*P-value* < 1.0×10^{-16}, 2.05 fold change) but similar by 2-wk-old. This result is consistent with work that demonstrated the presence of lymphatic capillaries throughout VEGFC and VEGFD in the muscles of humans and mice (Kivelä et al., 2007) and the role of VEGFC lymphatic regeneration in tissue repair of the intestinal muscle coat (Shimoda and Kato, 2006). For the PENK gene, a linear trend of expression can be identified, with lower ages having significantly higher expression (*P-value* < 1.0×10^{-6}) than advanced ages for the broiler line and at lower significance levels for the layer line. The level of expression in broilers was significantly higher than layers at 1-d-old (*P-value* < 4.0×10^{-4}, 1.61 fold change) but similar by 2-wk-old. This result is consistent with reports that during development, PENK mRNA is abundant in skeletal muscle, bone, and intestine, and that the levels of PENK mRNA tend to decrease as mice mature, with the exception of the brain, gut, lungs, and heart (Prasad et al., 2008).

Liver and Duodenum. From the analysis of gene expression profiles from experiment GSE15413, which compared the liver of chicken at two ages, only IGF1 was found over-expressed in 7-d-old relative to 0-d-old chickens (*P-value* < 2×10^{-4}, fold change 12.47) followed by prokineticin 2 (**PROK2**), which had the same profile but was only moderately differentially expressed. These results are consistent with reports that in avian species, IGF1 mRNA is found in the liver, muscle, kidney, testes, heart, ovary, brain, and intestine, and that the metabolic effects of IGF include increased amino acid and glucose uptake (Amills et al., 2003).

Also, Wang et al. (2007) reported that the hepatic expression of IGF1 was altered by hypothyroidism in chickens. As with heart and muscle, no neuropeptide gene differential expression was observed between the liver of female and male chicken embryos (study GSE6856; Itoh et al., 2007). Analysis of the GSE8016 study (Nakao et al., 2008) of liver from chicken and quail samples uncovered that PNOC, ADML, REL3, NMU, NPY, and ECRG4 were significantly under-expressed in chicken relative to quail (maximum *P-value* < 7.3×10^{-3}). From the analysis of experiment GSE15413, multiple neuropeptide genes were differentially expressed in the duodenum of newly hatched (0-d-old) relative to 7-d-old broiler chickens. The genes PENK and INS were significantly over-expressed in 0-d-old chickens (*P-value* < 9.0×10^{-4}, 6.23 fold change, and *P-value* < 3.5×10^{-3}, 14.72 fold change, respectively), and VEGFD was borderline over-expressed in 0-d-old chickens (*P-value* < 7.8×10^{-2}). The former finding is in agreement with studies showing that PENK mRNA is expressed in the human esophagus, gastrointestinal tract, pancreas, and gallbladder (Monstein et al., 2006) and enkephalins, such as PENK, have potent effects on gastrointestinal function, such as motility (Edin et al., 1980; Bitar and Makhlouf, 1982; Reynolds et al., 1984), intestinal secretion (Dobbins et al., 1980; McKay et al., 1981; Powell, 1981), and gastric acid secretion (Konturek et al., 1980; Feldman et al., 1980).

Brain and Hypothalamus. The limited number of neuropeptide genes differentially expressed in the brains of chickens under different conditions compared to other tissues and body parts was an unexpected finding. This result may be because the conditions compared within these studies did not allow the detection of major and consistent differential expression in the samples available. Experiment GSE12268 compared the brain of male and female embryos (6.5-d of incubation), and the present analysis determined that the pro-melanin-concentrating hormone

(**MCH**) gene was over-expressed; meanwhile, ECRG4 and GAST were borderline over-expressed in males (*P-value* < 2.5×10^{-2}, *P-value* < 9.5×10^{-2} and *P-value* < 9.5×10^{-2}, respectively). These results support the hypothesis that MCH acts as a neuromodulator involved in a wide variety of physiological and behavioral adaptations (arousal) with regard to feeding, drinking, and reproduction in birds (Cardot et al., 1999).

Results from the analysis of the other experiment (GSE6844; Itoh et al., 2007) that evaluated embryo female and male brains identified REL3 as being differentially over-expressed in females (*P-value* < 9.7×10^{-3}); meanwhile SCG1 and VIP were over-expressed in males (*P-value* < 3.3×10^{-2} and *P-value* < 2.1×10^{-2}). Also, the analysis of experiment GSE8693 detected REL3 as being over-expressed in the brain of females (*P-value* < 9.7×10^{-3}) relative to males. The observed profile of REL3 was expected because the relaxin hormone is renowned for its function in pregnancy, parturition, and other aspects of female reproduction (Agoulink, 2007). Relaxin-3 is a hypothalamic neuropeptide expressed in the nucleus incertus of the brainstem and plays a role in energy homeostasis (Tanaka et al., 2005; McGowan et al., 2009). Lastly, analysis of experiment GSE6868 (Rosenquist et al., 2007) that compared treated (homocysteine congenital defect cell culture) versus control neural crest samples did not uncover any differentially expressed neuropeptide genes. The lack of differential expression of the ADML gene in the brain is consistent with a report that the levels of ADML protein are almost undetectable in the chicken brain (Zudaire et al., 2005).

Oocyte and Gonads. The analysis of gene expression in oocytes across and within genetic lines from experiment GSE10231 uncovered that UTS2 was over-expressed in the genetic line with the long fertile period (DPF+) compared to the short fertile period (DPF-), high-growth (HG+),

and non-high-growth (HG-) genetic lines. This outcome is consistent with the role of UTS2 in reproduction that is associated with its spasmogenic activity and in humans, UTS2 has been linked to preeclampisa-eclampsia (Balat et al., 2005). Also, and as expected, numerous neuropeptide genes were differentially expressed between female and male gonads based on the profiles obtained from study GSE8693. Neuropeptide genes C-RF, SCG2, TKN1, PDGFA, PAHO, IGF1, and PDGFD were over-expressed in males relative to females (maximum *P-value* $< 4.0 \times 10^{-2}$ and 1.29 average fold change), and NEU2 was borderline under-expressed (*P-value* $< 6.0 \times 10^{-2}$). The neuropeptide genes with over-expression in females relative to males were ADML (*P-value* $< 3.5 \times 10^{-2}$ and 2.98 fold change) and GLUC (*P-value* $< 1.3 \times 10^{-2}$ and 2.84 fold change).

Other Tissues. The analysis of the gene expression information from experiment GSE8018 (Nakao et al., 2008), which investigated the effect of day length on quail using the chicken microarray platform, identified numerous differentially expressed neuropeptides with unique profiles. The expression of REL3 did not vary within the long-day cycle but was higher at 18 hr in the short-day cycle relative to the other time points in the same cycle (overall time-by-day cycle interaction *P-value* $< 2.8 \times 10^{-2}$). The expression of TSHB and cholecystokinin (**CCKN**) did not vary across the day within the day cycle but each was higher in the long-day cycle relative to the short-day cycle at every sampled time point (overall time-by-day cycle interaction *P-value* $< 5.8 \times 10^{-6}$ and *P-value* $< 1.4 \times 10^{-4}$, respectively). Lastly, no differential expression was found among the neuropeptide genes considered in the studies that compared neural crest cells treated with homocysteine versus control (study GSE6868), adipose tissue from lean and fat genetic lines (study GSE8010, Wang et al., 2007), circulating red and non-red blood cells (study

GSE9884, McIntyre et al., 2008), and histone H1 variants (study GSE8483, Takami and Nakayama, 1997).

Prohormone Cleavage Prediction

An outcome of the comprehensive survey of neuropeptide gene sequences is the ability to predict previously unidentified biologically active neuropeptides that can be used in high throughput experiments such as proteomic mass spectrometry experiments. We undertook the first prediction of cleavage sites in chicken prohormone sequences to gain insight into neuropeptide processing in avian species. The cleavage sites of all 24 chicken prohormone sequences, with empirically confirmed sequences and known or predicted neuropeptides available in UniProt (summarized in Table 1), were predicted using the empirically derived known motif and the human logistic regression cleavage models. The predictions were evaluated against the neuropeptides reported in UniProt. The comparison of the cleavage prediction models allows to assess the performance of the human cleavage model to predict avian cleavage sites.

The number of true positives (correctly predicted cleaved sites), true negatives (correctly predicted non-cleaved sites), false positives (incorrectly predicted cleaved sites), and false negatives (incorrectly predicted non-cleaved sites) obtained by the known motif and human models, respectively, were 36, 2811, 75, 13 and 35, 2851, 35, 14. The sensitivity, specificity, and correct classification rate of the known motif and human models, respectively, were 73.5, 97.4, 97.0 and 71.4, 98.8, and 98.3%. Model performance by individual neuropeptide prohormone is presented in Table 5. Overall, the sensitivity and the specificity of both models to predict cleavage was high, especially considering that neither model was trained using chicken

sequences. The highest number of correctly predicted cleavage sites was identified by the known motif model, and the highest number of correctly predicted non-cleavage sites was identified by the human model. Thus, the sensitivity (percentage of all cleaved sites that were correctly predicted) is higher in the known motif model and the specificity (percentage of all non-cleaved sites that were correctly predicted) is higher in the human model. Although the human model correctly predicted 40 additional non-cleaved sites and one less cleaved site, the overall difference in correct classification rate was minor (1.3%); this is because of the relatively higher number of non-cleaved sites than cleaved sites, which results in more weight for the correctly predicted cleaved sites relative to the correctly predicted non-cleaved sites. The previous results confirm that the processing of prohormone proteins into neuropeptides is similar between chicken and human species. Until more empirically confirmed neuropeptides are avilable to train and validate an avian model, the known motif and human models offer a good solution for predicting prohormone cleavages and determining the resulting neuropeptides in the chicken.

3.4 CONCLUSION

The role of neuropeptides on reproduction, development, growth, and health has been widely recognized. However, a comprehensive study of the representation and expression of neuropeptide genes in chicken has never been undertaken. In this study, the first survey of neuropeptide genes, prohormone sequences, and prohormone convertase enzyme genes in the chicken was completed. The integration of multiple bioinformatic resources allowed us to uncover evidence supporting 5 new neuropeptide genes, in addition to the62 previously reported in the chicken genome. Among the 62 chicken neuropeptide genes, 3 genes are not present in

Eutherian mammals. Due to insufficient evidence, 26 neuropeptide genes that are known in humans were unable to be detected in the chicken genome. A remarkable finding was that for most of the missing genes, another gene in the same neuropeptide family has been identified in the chicken genome. This finding suggests that neuropeptide genes have undergone less duplication, more gene loss, or both processes in the chicken than in the human. The high correct prediction of cleavage and non-cleavage sites in prohormones obtained with a model trained in human sequences indicates that the processing of prohormones into neuropeptides does not differ substantially between chicken and human species.

To gain a broad picture of the incidence of neuropeptide genes, we built a panel of expression across tissues and developmental stages. This panel will be of great value in streamlining neuropeptide research by helping to identify the tissues and developmental stages most likely to exhibit differential neuropeptide gene expression and subsequently, neuropeptide activity. Noteworthy findings include identifying the regions with the highest number of neuropeptide gene expression reports (brain, head, small intestine, and heart) and the most frequently reported expressed genes (PDGFA, SCG1, and ECRG4). To further understand the role of neuropeptides in reproduction, growth, and health, we analyzed the expression of neuropeptide genes across 22 microarray experiments that evaluated a wide range of ages, genders, tissues, genetic lines, and other conditions**.** Notable findings include various neuropeptide genes differentially expressed between the brain of male and female chickens; these include MCH, ECRG4, GAST, REL3, and SCG1. Also, numerous neuropeptide genes, including PNOC, PDGFA, PDGFB, PDGFD, ADML, ECRG4, PENK, and IGF1, were differentially expressed in the breast muscle between 7- and 0-d-old (just hatched) chickens. Lastly, the expression profiles of the neuropeptide genes

ADML, IGF1, VEGFC, VEGFD, and PENK across age differed significantly between broiler and layer genetic lines.

The chicken is a fundamental model for avian species, and more insight into the neuropeptide complement of this species can be expected from proteomic mass spectrometry studies in the chicken and also from the sequencing of the zebra finch, an avian model system used to study brain development, learning, and memory (The Zebra Finch Genome Consortium, 2005). The list of chicken neuropeptide genes will also support the annotation of homolog genes in avian species with genomes that are in the process of being sequenced and annotated or that do not have sequenced genomes. The series of bioinformatics steps used in this study is applicable to surveying neuropeptide or other gene sets in organisms with similar bioinformatics resources. The chicken neuropeptide gene sequences and prohormone cleavage prediction approaches are available at http://neuroproteomics.scs.uiuc.edu/neuropred.html. The expression panel developed here will facilitate neuropeptide research by aiding with identification of the tissues and developmental stages most likely to exhibit neuropeptide gene expression and subsequently, neuropeptide activity in avian species.

REFERENCES

Adams, M.D., Kelley J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R., Venter, J.C., 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252, 1651–1656.

Aikawa, S., Ishii, M., Yanagisawa, M., Sakakibara, Y., Sakurai, T., 2008. Effect of neuropeptide B on feeding behavior is influenced by endogenous corticotropin-releasing factor activities. Regul. Pept. 151, 147-152.

Alon, T., Hemo, I., Itin, A., Pe'er, J., Stone, J., Keshet, E., 1995. Vascular endothelial growth factor acts as a survival factor for newly formed retinal vessels and has implications for retinopathy of prematurity. Nat. Med. 1, 1024-1028.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403-410.

Álvaro, A., Martins, J., Araújo, I., Rosmaninho-Salgado, J., Ambrósio, A., Cavadas, C., 2008. Neuropeptide Y stimulates retinal neural cell proliferation involvement of nitric oxide. J. Neurochem. 105, 2501-2510.

Amare, A, Hummon, A.B., Southey, B.R., Zimmerman, T.A., Rodriguez-Zas, S.L., Sweedler, J.V., 2006. Bridging neuropeptidomics and genomics with bioinformatics, Prediction of mammalian neuropeptide prohormone processing. J. Proteome Res. 5, 1162-1167.

Amills, M., Jimenez, N., Villalba, D., Tor, M., Molina, E., Cubilo, D., Marcos, C. , Francesch, A. ,Sanchez, A., Estany, J., 2003. Identification of three single nucleotide polymorphisms in the chicken insulin-like growth factor 1 and 2 genes and their associations with growth and feeding traits. Poult. Sci. 82, 1485-1493.

Asnicar, M.A., Smith, D.P., Yang, D.D., Heiman, M.L., Fox, N., Chen, Y.F., Hsiung, H.M., Köster, A., 2001. Absence of cocaine -and amphetamine-regulated transcript results in obesity in mice fed a high caloric diet. Endocrinology 142, 4394-4400.

Baea, J.A., Parka, H.J., Seoa, Y.M., Rohb, J., Hsuehc, A.J.W., Chun, S.Y., 2008. Hormonal regulation of proprotein convertase subtilisin/kexin type 5 expression during ovarian follicle development in the rat. Mol. Cell. Endocrinol. 289, 29-37.

Bagnoli, P., Dal Monte M., Casini, G., 2003. Expression of neuropeptides and their receptors in the developing retina of mammals. Histol. Histopathol. 18, 1219-1242.

Balat, O., Aksoy, F., Kutlar, I., Ugur, M.G., Iyikosker, H., Balat, A., Anarat, R., 2005. Increased plasma levels of Urotensin-II in preeclampsia-eclampsia: a new mediator in pathogenesis? Eur. J. Obstet. Gynecol. Reprod. Biol. 120, 33-38.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. B. 57, 289-300.

Bennett, A.K., Hester, P.Y., Spurlock, D.E., 2006. Polymorphisms in vitamin D receptor, osteopontin, insulin-like growth factor 1 and insulin, and their associations with bone, egg and growth traits in a layer--broiler cross in chickens. Anim. Genet. 37, 283-286.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W. 2009. GenBank. Nucleic Acids Res. 37, D26-D31.

Birney E., Clamp M., Durbin R. 2004. GeneWise and Genomewise. Genome Res. 14:988-995.

Bitar, K. N., Makhlouf, G. M., 1982. Specific opiate receptors on isolated mammalian gastric smooth muscle cells. Nature 297, 72-74.

Burt, D.W., 2002. Origin and evolution of avian microchromosomes. Cytogenet. Genome Res. 96:97-112.

Burt, D.W., 2006. The Chicken Genome. Genome Dyn., 123-137.

Burt, D.W., 2007. Emergence of the chicken as a model organism: implications for agriculture and biology. Poult. Sci. 86, 1460-1471.

Cameron, M., Williams, H.E., Canane, A., 2004. Improved Gapped Alignment in BLAST. IEEE/ACM Trans. Comput. Biol. Bioinform.1:116-29.

Cardot, J., Griffond, B., Risold, P.Y., Blähser, S., Fellmann, D., 1999. Melanin-concentrating hormone-producing neurons in birds. J. Comp. Neurol. 411, 239-256.

Castro, A.A., Casagrande, T.S., Moretti, M., Constantino, L., Petronilho, F., Guerra, G.C. , Calo', G., Guerrini, R., Dal-Pizzol, F., Quevedo, J., Gavioli, E.C., 2009. Lithium attenuates behavioral and biochemical effects of neuropeptide S in mice. Peptides.

Chapman, A.M., Debski, E.A., 1995. Neuropeptide Y immunoreactivity of a projection from the lateral thalamic nucleus to the optic tectum of the leopard frog. Vis. Neurosci. 12, 1-9.

Cho, K., McFarlane, I.D., 1996. Physiological actions of the neuropeptide Antho-RNamide on antagonistic muscle systems in sea anemones. Neurosci. Lett. 219, 171-174.

Cogburn, L.A., Wang, X., Carre, W., Rejto, L., Porter, T.E., Aggrey, S.E., Simon, J., 2003. Systems-wide chicken DNA microarrays, gene expression profiling, and discovery of functional genes. Poult. Sci. 82, 939-951.

Cogburn, L.A., Porter, T.E., Duclos, M.J., Simon, J., Burgess, S.C., Zhu, J.J., Cheng, H.H., Dodgson, J.B., Burnside, J., 2007. Functional Genomics of the Chicken - A Model Organism. Poultry Science 86:2059-2094.

Conklin, D., Lofton-Day, C.E., Haldeman, B.A., Ching, A., Whitmore, T.E., Lok, S., Jaspers, S., 1999. Identification of INSL5, a new member of the insulin superfamily. Genomics 60, 50-66.

Cope, L.M., Irizarry, R.A., Jaffee, H.A., Wu, Z., Speed, T.P., 2004. A benchmark for Affymetrix GeneChip expression measures. Bioinformatics 20, 323-331.

D'Angelo, I., Brecha, N.C., 2004. Y2 receptor expression and inhibition of voltage-dependent Ca(2+) influx into rod bipolar cell terminals. Neuroscience 125, 1039-1049.

Dobbins, J., Racusen, L., Binder, H.J., 1980. Effect of d-alanine methionine enkephalin amide on ion transport in rabbit ileum. J. Clin. Invest. 66, 19-28.

Dodgson, J.B., Romanov, M.N., 2004. Use of Chicken Models for the Analysis of Human Disease. Current Protocols in Human Genetics, 15.5.1-15.5.12.

Duckert, P., Brunak, S., Blom, N., 2004. Prediction of proprotein convertase cleavage sites. Protein Engineering, Design & Selection, 107-112.

Duclos, M.J., 2005. Insulin-like growth factor-I (IGF-1) mRNA levels and chicken muscle growth. J. Physiol. Pharmacol. 56, 25-35.

Dun, S.L., Brailoiu, E., Wang, Y., Brailoiu, G.C., Liu-Chen, L.Y., Yang, J., Chang, J.K., Dun, N.J., 2006. Insulin-like peptide 5:expression in the mouse brain and mobilization of calcium. Endocrinology 147, 3243-3248.

Edin, R., Lundberg, J., Terenius, L., Dahlstrom, A., Hokfelt, T., Kewenter, J., Ahlman, H., 1980. Evidence for vagal enkephalinergic neural control of the feline pylorus and stomach. Gastroenterology 78, 492-497.

Ellegren, H., Hultin-Rosenberg, L., Brunström, B., Dencker, L., Kultima, K., Scholz, B., 2007. Faced with inequality: chicken do not have a general dosage compensation of sex-linked genes. BMC Biol. 5, 40.

Esposito, V., De Girolamo, P., Gargiulo, G., 1997. Neurotensin-like immunoreactivity in the brain of the chicken, Gallus domesticus. J. Anat. 191, 537-546.

Feldman, M., Walsh, J.H., Taylor, I.L., 1980. Effect of naloxone and morphine on gastric acid secretion and on serum gastrin and pancreatic polypeptide concentrations in humans. Gastroenterology 79, 294-298.

Fernández, A.P., Serrano, J., Tessarollo, L., Cuttitta, F., Martínez, A., 2008. Lack of adrenomedullin in the mouse brain results in behavioral changes, anxiety, and lower survival under stress conditions. Proc. Natl. Acad Sci. U.S.A. 105,12581-6.

Florin, S., Suaudeau, C., Meunier, J.C., Costentin, J., 1997. Orphan neuropeptide NocII, a putative pronociceptin maturation product, stimulates locomotion in mice. Neuroreport 8, 705-707.

Fricker, L.D., 2005. Neuropeptide-processing enzymes: applications for drug discovery. AAPS J. 7, E449-E455.

Fruttiger, M., Calver, A.R., Krüger, W.H., Mudhar, H.S., Michalovich, D., Takakura, N., Nishikawa, S., Richardson, W.D., 1996. PDGF mediates a neuron-astrocyte interaction in the developing retina. Neuron 17, 1117-1131.

Gyamera-Acheampong, C., Tantibhedhyangkul, J., Weerachatyanukul, W., Tadros, H., Xu, H., van de Loo, J.W., Pelletier, R.M., Tanphaichitr, N., Mbikay, M., 2006. Sperm from mice genetically deficient for the PCSK4 proteinase exhibit accelerated capacitation, precocious acrosome reaction, reduced binding to egg zona pellucida, and impaired fertilizing ability. Biol. Reprod. 74, 666-673.

Haugaard-Jönsson, L.M., Hossain, M.A., Daly, N.L., Craik, D.J., Wade, J.D., Rosengren, K.J., 2009. Structure of human insulin-like peptide 5 and characterization of conserved hydrogen bonds and electrostatic interactions within the relaxin framework. Biochem. J. 419, 619-627.

Hervieu, G., 2003. Melanin-concentration hormone functions in the nervous system: food intake and stress. Expert Opinion on Therapeutic Targets 7, 495-511.

Higuchi, K., Masaki, T., Gotoh, K., Chiba, S., Katsuragi, I., Tanaka, K., Kakuma T., Yoshimatsu, H., 2007. Apelin, an APJ Receptor Ligand, Regulates Body Adiposity and Favors the Messenger Ribonucleic Acid Expression of Uncoupling Proteins in Mice. Endocrinology 148, 2690-2697.

Hondo, M., Ishii, M., Sakurai, T., 2008. The NPB/NPW neuropeptide system and its role in regulating energy homeostasis, pain, and emotion. Results Probl. Cell Differ. 46, 239-256.

Hook, V., Funkelstein, L., Lu, D., Bark, S., Wegrzyn, J., Hwang, S.R., 2008. Proteases for processing proneuropeptides into peptide neurotransmitters and hormones. Annu. Rev. Pharmacol. Toxicol. 48, 393-423.

Hummon, A.B., Hummon, N.P., Corbin, R.W., Li, L., Vilim, F.S., Weiss, K.R., Sweedler, J.V., 2003. From precursor to final peptides: a statistical sequence-based approach to predicting prohormone processing. J. Proteome Res. 2, 650-656.

International Chicken Genome Sequencing Consortium. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432, 695-716.

Irizarry, R.A., Gautier, L., Bolstad, B.M. , Miller, C., Methods for Affymetrix Oligonucleotide Arrays. < http://bioconductor.org/packages/2.5/bioc/html/affy.html> November 10, 2009.

Itoh Y., Melamed, E., Yang, X., Kampf, K., Wang, S., Yehya, N., Van Nas, A., Replogle, K., Band, M.R., Clayton, D.F., Schadt, E.E., Lusis, A.J., Arnold, A.P., 2007. Dosage compensation is less effective in birds than in mammals. J. Biol. 6, 2.

Jiang N., Leach, L.J., Hu, X., Potokina, E., Jia, T., Druka, A., Waugh, R., Kearsey, M.J., Luo, Z.W. 2008. Methods for evaluating gene expression from Affymetrix mircoarray datasets. BMC Bioinformatics 9, 284.

Jozsa, R., Nemeth, J., Tamas, A., Hollosy, T., Lubics, A., Jakab, B., Olah, A., Lengvari, I., Arimura, A., Reglodi D., 2006. Short-term fasting differentially alters PACAP and VIP levels in the brains of rat and chicken. Ann. N.Y. Acad. Sci. 1070, 354-358.

Kivelä, R., Havas, E., Vihko, V., 2007. Localisation of lymphatic vessels and vascular endothelial growth factors-C and -D in human and mouse skeletal muscle with immunohistochemistry. Histochem. Cell Biol. 127, 31-40.

Konturek, S. J., Tasler, J., Cieszkowski, M., Mikos, E., Coy, D.H., Schally, A.V., 1980. Comparison of methionine-enkephalin and morphine in the stimulation of gastric acid secretion in the dog. Gastroenterology 78, 294-300.

Kubo, F., Nakagawa, S., 2009. Hairy1 acts as a node downstream of Wnt signaling to maintain retinal stem cell-like progenitor cells in the chick ciliary marginal zone. Development 136, 1823-1833.

Kuhar, M.J., Adams, S., Dominguez, G., Jaworski, J., Balkan, B., 2002. CART peptides. Neuropeptides 36, 1-8.

Ling, M.K., Hotta, E., Kilianova, Z., Haitina, T., Ringholm, A., Johansson, L., Gallo-Payet, N., Takeuchi, S., Schiöth, H.B. 2004. The melanocortin receptor subtypes in chicken have high preference to ACTH-derived peptides.Br. J. Pharmacol. 143,626-637.

Lukinius, A., Stridsberg, M., Wilander, E., 2003. Cellular expression and specific intragranular localization of chromogranin A, chromogranin B, and synaptophysin during ontogeny of pancreatic islet cells:an ultrastructural study. Pancreas 27, 38-46.

Maglott, D., Ostell, J., Pruitt K.D., Tatusova, T., 2005. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Research 33, D54-D58.

Martínez, A., Bengoechea, J.A., Cuttitta, F., 2006. Molecular evolution of proadrenomedullin N-terminal 20 peptide (PAMP): evidence for gene co-option. Endocrinology 147, 3457-3461.

McGlinn, A.M., Baldwin, D.A., Tobias, J.W., Budak, M.T., Khurana, T.S., Stone, R.A., 2007. Form-deprivation myopia in chick induces limited changes in retinal gene expression. Invest. Ophthalmol. Vis. Sci. 48, 3430-3436.

McGowan, B.M., Stanley, S.A., Ghatei, M.A., Bloom, S.R., 2009. Relaxin-3 and its role in neuroendocrine function. Ann. N.Y. Acad. Sci. 1160, 250-255.

McIntyre, B.A., Alev, C., Tarui, H., Jakt, L.M., Sheng, G., 2008. Expression profiling of circulating non-red blood cells in embryonic blood. BMC Dev. Biol. 8, 21.

McKay, J. S., Linaker, B.D., Turnberg, L.A., 1981. Influence of opiates on ion transport across rabbit ileal mucosa. Gastroenterology 80, 279-284.

McPherson, J.D., Dodgson, J., Krumlauf, R., Pourquié, O., Proposal to sequence the genome of the chicken. <http://www.genome.gov/Pages/Research/Sequencing /SeqProposals/Chicken_Genome.pdf> September 10, 2009.

Mirabeau, O., Perlas, E., Severini, C., Audero, E., Gascuel, O., Possenti, R., Birney, E., Rosenthal, N., Gross, C., 2007. Identification of novel peptide hormones in the human proteome by hidden Markov model screening. Genome Res. 17, 320-327.

Mollereau, C., Simons, M.J., Soularue, P., Liners, F., Vassart, G., Meunier, J.C., Parmentier, M., 1996. Structure, tissue distribution, and chromosomal localization of the prepronociceptin gene. Proc. Natl. Acad. Sci. U.S.A. 93, 8666-8670.

Monstein, H.J., Grahn, N., Ohlsson, B., 2006. Proenkephalin-A mRNA Is Widely Expressed in Tissues of the Human Gastrointestinal Tract. Eur. Surg. Res. 38, 464-468.

Montuenga, L.M., Martinez, A., Miller, M.J., Unsworth, E.J., Cuttitta, F., 1997. Expression of adrenomedullin and its receptor during embryogenesis suggests autocrine or paracrine modes of action. Endocrinology 138, 440-451.

Morash, M.G., MacDonald, A.B., Croll, R.P., Anini, Y., 2009. Molecular cloning, ontogeny and tissue distribution of zebrafish (Danio rerio) prohormone convertases: pcsk1 and pcsk2. General and Comparative Endocrinology 162,179-187.

Mori, Y., Ishiguro, H., Kuwabara, Y., Kimura, M., Mitsui, A., Kurehara, H., Mori, R., Tomado, K.,Ogawa, R., Katada, T., Harata, K., Fujii, Y., 2007. Expression of ECRG4 is an independent prognostic factor for poor survival in patients with esophageal squamous cell carcinoma. Oncol. Rep.18, 981-985.

Mudhar, H.S., Pollock, R.A., Wang, C., Stiles, C.D., Richardson, W.D., 1993. PDGF and its receptors in the developing rodent retina and optic nerve. Development 118, 539-552.

Nakao, N., Ono, H., Yamamura, T., Anraku, T., Takagi, T., Higashi, K., Yasuo, S., Katou, Y., Kageyama, S., Uno, Y., Kasukawa, T., Iigo, M., Sharp, P.J., Iwasawa, A., Suzuki, Y., Sugano, S., Niimi, T., Mizutani, M., Namikawa, T., Ebihara, S., Ueda, H.R., Yoshimura, T., 2008. Thyrotrophin in the pars tuberalis triggers photoperiodic response. Nature 452, 317-322.

NCBI. 2002. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information <http://www.ncbi.nlm.nih.gov/> October 10, 2009.

Pape, H.C., Jüngling, K., Seidenbecher, T., Lesting, J., Reinscheid, R.K., 2009. Neuropeptide S: A transmitter system in the brain regulating fear and anxiety. Neuropharmacology. June 10.

Pierce, E.A., Foley, E.D., Smith, L.E., 1996. Regulation of vascular endothelial growth factor by oxygen in a model of retinopathy of prematurity. Arch. Ophthalmol. 114, 1219-1228.

Powell, D. W., 1981. Muscle or mucosa:the site of action of antidiarrheal opiates? Gastroenterology 80, 406-408.

Prasad, S.K., Clerk, A., Cullingford, T.E., Chen, A.W., Kemp, T.J., Cannell, T.M., Cowie, M.R., Petrou, M., 2008. Gene expression profiling of human hibernating myocardium:increased expression of B-type natriuretic peptide and proenkephalin in hypocontracile vs normally-contracting regions of the heart. Eur. J. Heart Fail. 10, 1177-1180.

Provis, J.M., Leech, J., Diaz, C.M., Penfold, P.L., Stone, J., Keshet, E., 1997. Development of the human retinal vasculature:cellular relations and VEGF expression. Exp. Eye Res. 65, 555-568.

Reynolds, J. C., Ouyang, A., Cohen, S., 1984. Evidence for an opiate-mediated pyloric sphincter reflex. Am. J. Physiol. 246, G130-136.

Richards M.P., McMurtry J.P. 2008. Expression of proglucagon and proglucagon-derived peptide hormone receptor genes in the chicken. Gen. Comp. Endocrinol. 156,323-338.

Rizzolo, L.J., Chen, X., Weitzman, M., Sun, R., Zhang, H., 2007. Analysis of the RPE transcriptome reveals dynamic changes during the development of the outer blood-retinal barrier. Mol. Vis. 13, 1259-1273.

Rosenquist, T.H., Bennett, G.D., Brauer, P.R., Stewart, M.L., Chaudoin, T.R., Finnell, R.H., 2007. Microarray analysis of homocysteine-responsive genes in cardiac neural crest cells in vitro. Dev. Dyn. 236, 1044-1054.

Sayers, E.W., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S., Feolo, M., Geer, L.Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D.J., Madden, T.L., Maglott, D.R., Miller, V., Mizrachi, I., Ostell, J., Pruitt, K.D., Schuler, G.D., Sequeira, E., Sherry, S.T., Shumway, M., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusova, T.A., Wagner, L., Yaschenko, E., Ye, J., 2009. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 37, D5-15.

Schippert, R., Schaeffel, F., Feldkaemper, M.P., 2008. Microarray analysis of retinal gene expression in chicks during imposed myopic defocus. Mol. Vis. 14, 1589-1599.

Schwippert, W.W., Röttgen, A., Ewert., J.P., 1998. Neuropeptide Y (NPY) or fragment NPY 13-36, but not NPY 18-36, inhibit retinotectal transfer in cane toads Bufo marinus. Neurosci. Lett. 253, 33-36.

Shimoda, H., Kato, S., 2006. A model for lymphatic regeneration in tissue repair of the intestinal muscle coat. Int. Rev. Cytol. 250, 73-108.

Smith, B. J., Ko, Y., Southey, B.R., Rodriguez-Zas, S.L., 2007. BEEHIVE - A suite of tools to manage, analyze and interpret honey bee microarray experiments. Workshop on honey bee genomics & biology. Cold Spring Harbor Laboratory, May 6 - 8, 2007. Cold Spring Harbor, NY.

Southey, B.R., Amare, A., Zimmerman, T.A., Rodriguez-Zas, S.L., Sweedler, J.V., 2006a. NeuroPred: a tool to predict cleavage sites in neuropeptide precursors and provide the masses of the resulting peptides. Nucleic Acids Res. 34, W267-W272.

Southey, B.R., Rodriguez-Zas, S.L., Sweedler, J.V. 2006b. Prediction of neuropeptide prohormone cleavages with application to RFamides. Peptides 27, 1087-1098.

Southey, B.R., Sweedler, J.V., Rodriguez-Zas, S.L., 2008a. A python analytical pipeline to identify prohormone precursors and predict prohormone cleavage sites. Front. Neuroinformatics, 2:7.

Southey, B.R., Sweedler, J.V., Rodriguez-Zas, S.L., 2008b. Prediction of neuropeptide cleavage sites in insects. Bioinformatics 24, 815-825.

Southey, B.R., Rodriguez-Zas, S.L., Sweedler, J.V., 2009. Characterization of the prohormone complement in cattle using genomic libraries and cleavage prediction approaches. BMC Genomics 10, 228.

Speed, T.P., McPeek, M.S., Evans, S.N., 1992. Robustness of the no-interference model for ordering genetic markers. Proc. Natl. Acad. Sci. U. S. A. 89, 3103-3106.

Stekel, D., 2003. Microarray Bioinformatics. Cambridge University Press, New York.

Stern, C.D., 2005. The chick; a great model system becomes even greater. Dev. Cell 8, 9-17.

Stone, J., Itin, A., Alon, T., Pe'er, J., Gnessin, H., Chan-Ling, T., Keshet. E., 1995. Development of retinal vasculature is mediated by hypoxia-induced vascular endothelial growth factor (VEGF) expression by neuroglia. J. Neurosci. 15, 4738-4747.

Strand, F.L., 1999. Neuropeptides. Massachusetts Institute of Technology, Cambridge, Massachusetts.

Takami, Y., Nakayama, T., 1997. A single copy of linker H1 genes is enough for proliferation of the DT40 chicken B cell line, and linker H1 variants participate in regulation of gene expression. Genes Cells 2, 711-723.

Tanaka, M., Iijima, N., Miyamoto, Y., et al., 2005. Neurons expressing relaxin 3/INSL 7 in the nucleus incertus respond to stress. Eur. J. Neurosci. 21, 1659-1670.

Tegge, A.N., Southey, B.R., Sweedler, J.V., Rodriguez-Zas, S.L., 2008. Comparative analysis of neuropeptide cleavage sites in human, mouse, rat, and cattle. Mamm. Genome 19, 106-120.

The UniProt Consortium, 2007. The Universal Protein Resource (UniProt). Nucleic Acids Research 35, D193-D197.

The Zebra Finch Genome Consortium. 2005. Proposal to Sequence the Genome of the Zebra Finch (Taeniopygia guttata).<http://www.songbirdgenome.org/pdfs/ ZebraFinchGenome NHGRIjuly05a.pdf > November 10, 2009.

Udono, T., Takahashi, K., Nakayama, M., Murakami, O., Durlu, Y.K., Tamai, M.,Shibahara, S., 2000. Adrenomedullin in cultured human retinal pigment epithelial cells. Invest. Ophthalmol. Vis. Sci. 41, 1962-1970.

Van Den Akker, N.M., Lie-Venema, H., Maas, S., Eralp, I., DeRuiter, M.C., Poelmann, R.E., Gittenberger-De Groot, A.C., 2005. Platelet-derived growth factors in the developing avian heart and maturating coronary vasculature. Dev. Dyn. 233, 1579-1588.

Wang, G., Yan, B., Deng, X., Li, C., Hu, X., Li, N., 2005. Insulin-like growth factor 2 as a candidate gene influencing growth and carcass traits and its bialleleic expression in chicken. Sci. China C. Life Sci. 48, 187-194.

Wang, H.B., Li, H., Wang, Q.G., Zhang, X.Y., Wang, S.Z., Wang, Y.X., Wang, X.P., 2007. Profiling of chicken adipose tissue gene expression by genome array. BMC Genomics 8, 193.

Xiao, P.F., He, N.Y., Liu, Z.C., He, Q.G., Sun, X., Lu, Z.H., 2005. In situ synthesis of oligonucleotide arrays by using soft lithography. Nucleic Acids Res. 33(8), e75.

Yang, X., Chrisman, H., Weijer, C.J., 2008. PDGF signalling controls the migration of mesoderm cells during chick gastrulation by regulating N-cadherin expression. Development 135, 3521-3530.

Zheng, Q., Zhang, Y., Chen, Y., Yang, N., Wang, X., Zhu. D., 2009. Systematic identification of genes involved in divergent skeletal muscle growth rates of broiler and layer chickens. BMC Genomics 10, 87.

Zhou, H., Mitchell, A.D., McMurtry, J.P., Ashwell, C.M., Lamont, S.J., 2005. Insulin-like growth factor-I gene polymorphism associations with growth, body composition, skeleton integrity, and metabolic traits in chickens. Poult. Sci. 84, 212-219.

Zudaire, E., Cuesta, N., Martínez, A., Cuttitta, F., 2005. Characterization of adrenomedullin in birds. Gen. Comp. Endocrinol. 143, 10-20.

Table 1. Neuropeptide and convertase gene and protein master list

Neuropeptide Prohormone					
Abbreviated Name	**Name**	**UniProtID[1]**	**GeneID[2]**	**UniGeneID[3]**	**Evidence in Chicken[4]**
ADM2	Intermedin	NA[5]	NA	NA	NA
ADML	Adrenomedullin	NA	423042	Gga.12006	NA
ANF	Atrial natriuretic factor	P18908	395765	Gga.5157	protein
ANFB	Natriuretic peptides B	NA	NA	NA	NA
ANFC	C-type natriuretic peptide	A9CDT6	419487	Gga.12392	transcript
APEL	Apelin	NA	NA	NA	NA
C-RF AMIDE	C-RF amide peptide	B0LF68	420716	Gga.3202	predicted
CALC/CALCA	Calcitonin/Calcitonin gene-related peptide 1	P07660, P10286	396256	Gga.4991	transcript/ protein
CALCB	Calcitonin gene-related peptide 2	NA	NA	NA	NA
CART	Cocaine- and amphetamine-regulated transcript protein	NA	NA	NA	NA
CCKN	Cholecystokinin	Q9PU41	414884	Gga.2441	protein
CMGA	Chromogranin-A	NA	423420	Gga.19002	NA
CNP1	C-type natriuretic peptide 1	A9CDT5	NA	Gga.47230	transcript
COLI	Pro-opiomelanocortin	Q9YI93	422011	Gga.6271	predicted
CORT	Cortistatin	NA	NA	NA	NA
CRF	Corticoliberin	Q703P0	404297	Gga.11323	transcript
ECRG4	Augurin (Esophageal cancer-related gene 4 protein)	NA	771055	Gga.8435	NA
EDN1	Endothelin-1	NA	420854	Gga.25090	NA
EDN2	Endothelin-2	NA	419559	Gga.8238	NA
EDN3	Endothelin-3	Q3MU75	768509	Gga.22840	transcript
GALA	Galanin	P30802	423117	Gga.12649	protein
GALP	Galanin-like peptide	NA	NA	NA	NA
GAST	Gastrin	P09859	396365	Gga.782	protein
GHRL	Obestatin	Q8AV73, Q7T2V1	408185	Gga.16	homology
GIP	Gastric inhibitory polypeptide	A1DPK0	419989	Gga.7981	transcript
GLUC	Glucagon	P68259	396196	Gga.704	protein
GON1	Progonadoliberin-1	P37042	770134	Gga.41802	protein
GRP	Gastrin-releasing peptide	P01295	425213	Gga.43422	protein
HEPC	Hepcidin	NA	NA	NA	NA
IAPP	Islet amyloid polypeptide	Q90743	396362	Gga.780	transcript

Table 1 (cont.)

Neuropeptide Prohormone					
Abbreviated Name	**Name**	**UniProtID[1]**	**GeneID[2]**	**UniGeneID[3]**	**Evidence in Chicken[4]**
IGF1	Insulin-like growth factor I	P18254	418090	Gga.850	protein
IGF2	Insulin-like growth factor 2 (somatomedin A)	P33717	395097	Gga.8511	protein
INS	Insulin	P67970	396145	Gga.673	protein
INSL3	Insulin-like 3	NA	NA	NA	NA
INSL5	Insulin-like 5	NA	NA	NA	NA
INSL6	Insulin-like 6	NA	NA	NA	NA
KISS1	Metastasis-suppressor KiSS-1	NA	NA	NA	NA
MCH	Pro-melanin-concentrating hormone	NA	418091	Gga.14659	NA
MOTI	Motilin	Q9PRP6	768422	NA	protein
NEU1	Oxytocin	Q2ACD0	768516	NA	predicted
NEU2	Neurophysin-II	P24787	396101	Gga.652	transcript
NEUT	Neurotensin	P13724	417883	Gga.10167	protein
NMB	Neuromedin-B	A0MAR5	415333	Gga.8071	transcript
NMS	Neuromedin-S	NA	NA	NA	NA
NMU	Neuromedin-U	P34963	422748	Gga.18392	protein
NPB	Neuropeptide B	NA	NA	NA	NA
NPFF	Neuropeptide FF	NA	NA	NA	NA
NPS	Neuropeptide S	NA	NA	NA	NA
NPW	Neuropeptide W	NA	NA	NA	NA
NPY	Neuropeptide Y	P28673	396464	Gga.837	homology
OREX	Orexin	Q8AV17	374005	Gga.11	transcript
OSTN	Osteocrin (Musclin)	A5JNH0	424907	Ggta.13448	transcript
OX26	Orexigenic neuropeptide QRFP	B2CL09	771867	NA	transcript
PACA	Pituitary adenylate cyclase-activating polypeptide	P41534	408251	Gga.616	protein
PAHO	Pancreatic polypeptide	P68248	395564	Gga.308	protein
PCSK1N	Proprotein convertase subtilisin/kexin type 1 inhibitor	NA	NA	NA	NA
PDGFA	Platelet-derived growth factor alpha polypeptide	Q90WK2, Q9PUF7	374196	Gga.3899	transcript
PDGFB	Platelet-derived growth factor beta polypeptide	Q90W23	374128	Gga.71	transcript
PDGFD	Platelet derived growth factor D	O57658	418978	Gga.43662	transcript

Table 1 (cont.)

Neuropeptide Prohormone					
Abbreviated Name	**Name**	**UniProtID[1]**	**GeneID[2]**	**UniGeneID[3]**	**Evidence in Chicken[4]**
PDYN	Proenkephalin-B	NA	NA	NA	NA
PENK	Proenkephalin	NA	421131	Gga.11430	NA
PNOC	Prepronociceptin	NA	422019	Gga.10041	NA
PROK2	Prokineticin 2	NA	771674	Gga.10528	NA
PRRP	Prolactin-releasing peptide	A3RJ26	424018	Gga.10552	predicted
PTHR	Parathyroid hormone-related protein	P17251	396281	Gga.2626	protein
PTHY	Parathyroid hormone	P15743	396436	Gga.78	homology
PYY	Peptide YY	P29203	NA	NA	protein
PYY2	Putative peptide YY-2	NA	NA	NA	NA
REL1	pro-relaxin 1	NA	NA	NA	NA
REL2	pro-relaxin 2	NA	NA	NA	NA
REL3	Relaxin-3	B1AC67	427223	Gga.37019	transcript
RES18	Regulated endocrine-specific protein 18	NA	NA	NA	NA
RFRP	Neuropeptide VF precursor	Q6T2D1, Q75XU6	378785	Gga.9285	transcript
RNP	Renal natriuretic peptide	A9CDT7	NA	NA	transcript
SCG1	Secretogranin-1	NA	421312	Gga.10025	NA
SCG2	Secretogranin-2	NA	424808	Gga.11999	NA
SECR	Secretin	P01280	423015	Gga.14227	protein
SLIB	Somatoliberin	Q1KNA8, Q1KNA7	419178	Gga.11231	transcript
SMS	Somatostatin	P33094	396279	Gga.742	homology
SPXN	Spexin	NA	NA	NA	NA
TAC4/TKN4	Tachykinin-4	NA	NA	NA	NA
TIP39	Parathyroid hormone 2	NA	NA	NA	NA
TKN1	Tachykinin, precursor 1	NA	420573	Gga.12286	NA
TKNK	Tachykinin 3	NA	NA	NA	NA
TOR2X	Torsin family 2, member A	NA	NA	NA	NA
TRH	Prothyroliberin	Q6ZXC3	414344	Gga.19489	transcript
TSHB	Thyroid-stimulating hormone subunit beta	O57340	395937	Gga.551	transcript
UCN1	Urocortin	NA	NA	NA	NA
UCN2	Urocortin 2	NA	NA	NA	NA
UCN3	Urocortin 3	NA	769274	Gga.11141	NA
UTS2	Urotensin 2	Q6Q2J6	404535	Gga.14388	transcript

Table 1 (cont.)

Neuropeptide Prohormone					
Abbreviated Name	**Name**	**UniProtID[1]**	**GeneID[2]**	**UniGeneID[3]**	**Evidence in Chicken[4]**
UTS2D	Urotensin II-related peptide	Q6Q273	404534	Gga.9482	transcript
VEGFC	Vascular endothelial growth factor C	NA	422573	Gga.12347	NA
VEGFD	Vascular endothelial growth factor D	Q8QGD7	395255	Gga.3219	transcript
VIP	Vasoactive intestinal peptides	P48143	396323	Gga.666	protein
Prohormone Convertase Enzyme					
PCSK1	Proprotein convertase subtilisin/kexin type 1	NA	395137	Gga.9357	NA
PCSK2	Proprotein convertase subtilisin/kexin type 2	NA	395136	Gga.9404	NA
PCSK3	Furin	Q91000	395457	Gga.1751	transcript
PCSK4	Proprotein convertase subtilisin/kexin type 4	NA	NA	NA	NA
PCSK5	Proprotein convertase subtilisin/kexin type 5	NA	395456	Gga.12660	NA
PCSK6	Proprotein convertase subtilisin/kexin type 6	NA	395454	Gga.21090	NA
PCSK7	Proprotein convertase subtilisin/kexin type 7	Q5ZKB5	395455	Gga.5311	transcript

[1]: UniProt identifier
[2]: Gene database identifier
[3]: UniGene database identifier
[4]: Evidence in chicken in UniProt at the protein or transcript level, inferred from homology or predicted
[5]: NA= Not available

Table 2. Abbreviated distribution of neuropeptide and convertase gene EST across tissues and stages

Neuropeptide Prohormones	Unigene ID[1]	Brain	Head	Ovary	Small Intestine	Embryo Stage	Adult Stage
ADML	Gga.12006	1[2]	0	0	0	1	1
ANF	Gga.5157	1	0	0	0	1	1
ANFC	Gga.12392	1	0	0	0	0	1
C-RF AMIDE	Gga.3202	1	1	1	0	1	1
CALCA	Gga.4991	1	1	0	0	1	0
CCKN	Gga.2441	1	0	1	1	1	1
CMGA	Gga.19002	1	1	1	1	1	1
COLI	Gga.6271	1	0	0	0	0	0
ECRG4	Gga.8435	1	1	1	1	1	1
EDN1	Gga.25090	0	0	0	0	1	1
EDN2	Gga.8238	1	0	1	1	1	1
EDN3	Gga.22840	0	0	1	0	1	1
GALA	Gga.12649	0	0	0	1	0	1
GHRL	Gga.16	0	0	1	0	1	1
GIP	Gga.7981	0	0	0	1	0	1
GLUC	Gga.704	0	1	0	1	1	1
GRP	Gga.43422	0	0	0	0	0	0
IAPP	Gga.780	1	1	0	0	1	1
IGF1	Gga.850	0	0	1	0	0	1
IGF2	Gga.8511	0	0	0	0	0	1
INS	Gga.673	0	0	1	0	0	1
MCH	Gga.14659	1	0	0	0	1	0
NEU2	Gga.652	1	0	0	0	1	1
NEUT	Gga.10167	1	1	0	1	1	1
NMB	Gga.8071	1	0	0	0	1	1
NMU	Gga.18392	1	0	0	1	1	1
NPY	Gga.837	1	0	1	0	1	1
OREX	Gga.11	0	0	0	0	1	0
OSTN	Gga.13448	0	1	0	0	1	1
PACA	Gga.616	1	1	0	0	1	1
PDGFA	Gga.3899	1	1	1	1	1	1
PDGFB	Gga.71	1	1	1	0	1	1
PDGFD	Gga.43662	1	1	0	0	1	1
PENK	Gga.11430	1	1	1	0	1	1
PNOC	Gga.10041	1	0	0	0	1	0
PROK2	Gga.10528	1	0	1	0	1	1
PRRP	Gga.10552	0	0	1	0	0	1
PTHR	Gga.2626	0	0	0	0	1	0
REL3	Gga.37019	1	0	1	0	1	1
RFRP	Gga.9285	1	1	0	0	1	1

Table 2 (cont.)

Neuropeptide Prohormones	Unigene ID[1]	Brain	Head	Ovary	Small Intestine	Embryo Stage	Adult Stage
SCG1	Gga.10025	1	1	1	0	1	1
SCG2	Gga.11999	1	1	0	0	1	1
SECR	Gga.14227	0	0	0	1	0	1
SLIB	Gga.11231	1	1	0	0	1	0
SMS	Gga.742	1	0	0	1	1	1
TKN1	Gga.12286	1	0	0	1	0	1
UCN3	Gga.11141	0	0	0	1	0	1
UTS2D	Gga.9482	0	1	0	0	1	0
VEGFC	Gga.12347	0	1	0	0	1	1
VEGFD	Gga.3219	1	1	1	1	1	1
VIP	Gga.666	1	1	0	1	1	1
Total		**33**	**7**	**2**	**6**	**8**	**9**
Prohormone Convertases							
PCSK1	Gga.9357	N/A	N/A	N/A	N/A	N/A	N/A
PCSK1 similar	Gga.31439	1	0	0	1	1	1
PCSK2	Gga.9404	1	1	0	1	1	1
PCSK3	Gga.1751	1	0	0	0	0	1
PCSK5	Gga.12660	1	0	1	1	1	1
PCSK6	Gga.21090	0	0	1	1	1	1
PCSK7	Gga.5311	1	1	1	1	1	1
Total		**5**	**2**	**3**	**5**	**5**	**6**

[1]: UniGene database identifier
[2]: 1 denotes presence and 0 denotes absence

Table 3. Abreviated description of the 22 chicken microarray experiments analyzed

Exp. ID[1]	Tissue	Gender	Age	Reference
GSE6543	Retina	Female	1-wk	McGlinn et al.. (2007)
GSE6843	Embryonic heart	M/F[2]	Late stage embryo	Itoh et al. (2007)
GSE6844	Embryonic brain	M/F	Late stage embryo	Itoh et al. (2007)
GSE6856	Embryonic liver	M/F	Late stage embryo	Itoh et al. (2007)
GSE6868	Neural tube explants	NA[3]	Stage 9+ embryo	Rosenquist et al. (2007)
GSE7176	Retinal epithelium	NA	7-d-old embryo	Rizzolo et al. (2007)
GSE8010	Adipose Tissue	Female	7-wk	Wang et al. (2007)
GSE8016	Liver	Female	NA	Nakao et al. (2008)
GSE8018	Hypothalamus	NA	NA	Nakao et al. (2008)
GSE8483	DT40 cells	NA	NA	Takami et al. (1997)
GSE8693	Embryonic heart	M/F	18-d-old embryo	Ellegren et al. (2007)
GSE8693	Embryonic brain	M/F	18-d-old embryo	Ellegren et al. (2007)
GSE8693	Embryonic gonad	M/F	18-d-old embryo	Ellegren et al. (2007)
GSE9251	Pectoralis muscles	Female	1-d-old to 8-wk-old	Zheng et al. (2009)
GSE9884	Embryonic heart blood	NA	Embryo	McIntyre et al. (2008)
GSE10231	F1 oocyte stage	Female	NA	NA
GSE11439	Retina	Male	9-d-old	Schippert et al. (2008)
GSE12268	Brain	M/F	Stage 29 embryo	NA
GSE15382	Retina	NA	Embryo	Kubo et al. (2009)
GSE15413	Liver	NA	Newly hatched/7-d-old	NA
GSE15413	Brain	NA	Newly hatched/7-d-old	NA
GSE15413	Duodenum	NA	Newly hatched/7-d-old	NA

[1]: ID = Identifer

[2]: M/F = Male and Female

[3]: NA= Not available

Table 4. Number of differentially expressed neuropeptide and convertase genes (*P-value* < 0.005) across 22 microarray studies grouped by tissue type

Prohormone	UniGene Probe[1]	Retina	Heart-breast	Brain-head	Liver-duod.[2]	Oocyte-gonad	Others	Total by Probe
ADML	**Gga.12006.1.S1_at**	1	2	0	0	1	1	5
ANF	**Gga.5157.1.S1_at**	0	0	0	0	0	0	0
ANFC	**Gga.12392.1.S1_a_at**	0	1	0	0	0	1	2
CALCA	**Gga.4991.2.S1_a_at**	1	0	0	0	0	0	1
CCKN	**GgaAffx.21834.1.S1_s_at**	0	0	0	0	0	1	1
CCKN	**Gga.2441.1.S1_at**	0	0	0	0	0	0	0
CMGA	**GgaAffx.21576.1.S1_s_at**	1	1	0	1	0	1	4
CMGA	**Gga.12437.2.S1_at**	0	1	0	0	0	0	1
COLI	**Gga.6271.1.S1_at**	0	1	0	0	0	1	2
CRF	**Gga.11323.1.S1_at**	0	0	0	0	1	0	1
ECRG4	**Gga.8435.1.S1_at**	0	2	0	0	0	1	3
ECRG4	**Gga.11232.1.A1_at**	0	1	0	0	0	0	1
GALA	**Gga.12649.1.S1_at**	0	0	0	0	0	0	0
GAST	**Gga.782.1.S1_at**	0	1	0	0	0	0	1
GHRL	**Gga.16.1.S1_at**	0	0	0	0	0	0	0
GIP	**Gga.7981.1.S1_at**	0	1	0	0	0	0	1
GLUC	**GgaAffx.21780.2.S1_s_at**	0	0	0	0	1	0	1
IAPP	**Gga.780.1.S1_at**	0	1	0	0	0	0	1
IGF1	**Gga.850.1.S1_at**	0	2	0	1	1	0	4
IGF2	**Gga.8511.1.S1_at**	0	1	0	0	0	1	2
INS	**Gga.673.1.S1_at**	0	0	0	1	0	1	2
MCH	**Gga.14659.1.S1_at**	0	0	1	0	0	0	1
NEU2	**Gga.652.1.S1_at**	0	0	0	0	0	0	0
NEUT	**Gga.10167.1.S1_at**	1	0	0	0	0	0	1
NMB	**Gga.8071.1.S1_a_at**	0	0	0	0	0	0	0
NMU	**Gga.18392.1.S1_at**	0	1	0	1	0	1	3
NPY	**Gga.837.1.S1_a_at**	0	0	0	0	0	1	1
OREX	**Gga.11.1.S1_at**	0	0	0	0	0	0	0
OSTN	**Gga.13448.1.S1_at**	0	1	0	0	0	0	1
PACA	**Gga.11409.1.S1_at**	0	1	0	0	0	0	1
PACA	**Gga.616.1.S1_s_at**	0	0	0	0	0	0	0
PAHO	**Gga.308.1.S1_at**	0	0	0	0	1	0	1
PDGFA	**Gga.3899.3.S1_a_at**	1	1	0	0	1	0	3
PDGFB	**Gga.71.1.S1_at**	0	1	0	0	0	0	1
PDGFD	**Gga.9675.1.S1_at**	0	2	0	0	1	0	3
PENK	**Gga.11430.1.S1_at**	0	2	0	1	0	0	3
PNOC	**Gga.10041.1.S1_a_at**	0	2	0	0	0	0	2
PNOC	**Gga.10041.2.A1_at**	0	1	0	0	0	1	2
PNOC	**GgaAffx.20191.1.S1_s_at**	0	0	0	0	0	0	0

Table 4 (cont.)

Prohormone	UniGene Probe[1]	Retina	Heart-breast	Brain-head	Liver-duod.[2]	Oocyte-gonad	Others	Total by Probe
PROK2	**Gga.10528.1.S1_a_at**	0	0	0	1	0	0	1
PROK2	**Gga.10528.2.A1_at**	0	0	0	0	0	0	0
PRRP	**Gga.10552.1.S1_at**	0	1	0	0	0	0	1
PTHR	**Gga.2626.1.S1_at**	0	1	0	0	0	1	2
PTHY	**Gga.78.1.A1_at**	0	0	0	0	0	0	0
REL3	**Gga.12454.1.S1_at**	0	0	1	0	0	2	3
RFRP	**Gga.9285.1.S1_at**	0	0	0	0	0	1	1
SCG1	**Gga.10025.1.S1_at**	0	0	1	0	0	0	1
SCG2	**Gga.11999.1.S1_at**	0	0	0	0	0	0	0
SCG2	**Gga.11999.1.A1_s_at**	0	1	0	0	1	0	2
SCG2	**Gga.11999.1.A1_at**	0	0	0	0	0	0	0
SECR	**Gga.14227.1.S1_at**	0	0	0	0	0	0	0
SLIB	**Gga.11231.1.S1_at**	0	0	0	0	0	0	0
SMS	**Gga.742.1.S1_at**	0	1	0	0	0	0	1
TKN1	**Gga.12286.1.S1_at**	0	1	0	0	1	0	2
TRH	**Gga.19489.1.A1_at**	0	1	0	0	0	0	1
TRH	**Gga.19489.1.S1_at**	0	0	0	0	0	0	0
TSHB	**Gga.551.1.S1_at**	0	1	0	0	0	1	2
UCN3	**Gga.11141.1.S1_at**	0	0	0	0	0	0	0
UTS2	**Gga.14388.1.S1_at**	0	0	0	0	0	0	0
UTS2D	**Gga.9482.1.S1_at**	2	0	0	0	0	0	2
VEGFC	**Gga.12347.1.S1_at**	1	1	0	0	0	0	2
VEGFD	**Gga.3219.1.S1_at**	1	1	0	0	0	0	2
VIP	**Gga.666.1.S1_a_at**	1	0	1	0	0	0	2
Total by Tissue		**10**	**36**	**4**	**6**	**9**	**16**	**81**

Table 4 (cont.)

Prohormone Convertases	UniGene Probe[1]	Retina	Heart-breast	Brain-head	Liver-duod.[2]	Oocyte-gonad	Others	Total by Probe
PCSK2	**Gga.2786.1.S1_at**	0	1	0	0	0	0	1
PCSK2	**Gga.9404.1.S1_at**	0	0	0	0	0	0	0
PCSK3	**Gga.1751.1.S1_at**	0	0	0	0	1	0	1
PCSK5	**Gga.247.1.S1_at**	1	0	0	0	0	0	1
PCSK5	**Gga.12660.2.S1_a_at**	0	0	0	0	0	1	1
PCSK6	**Gga.20041.1.S1_at**	0	1	0	1	1	0	3
PCSK6	**GgaAffx.20832.1.S1_s_at**	0	0	0	0	1	0	1
PCSK6	**Gga.246.1.S1_at**	0	0	0	0	0	0	0
PCSK7	**GgaAffx.12272.1.S1_s_at**	0	0	0	0	0	0	0
PCSK7	**Gga.17539.1.S1_s_at**	0	0	0	0	1	0	1
Total by Tissue		**1**	**2**	**0**	**1**	**4**	**1**	**9**

1: UniGene identifier of microarray gene probe
2: Duodenum

Table 5. Evaluation of the prediction of cleavage sites in chicken prohormone sequences

	Known Motif Model				Human Model			
Neuropeptide Prohormone	**TP**	**TN**	**FP**	**FN**	**TP**	**TN**	**FP**	**FN**
ANF	0	107	4	1	0	107	4	1
CALC	2	105	2	0	2	107	0	0
CALCA	2	91	3	0	1	93	1	1
CCKN	2	102	0	2	2	102	0	2
GALA	2	88	0	0	1	88	0	1
GLUC	5	171	2	2	5	173	0	2
GON1	1	63	1	0	1	64	0	0
GRP	1	104	0	1	2	103	1	0
IGF1	0	96	4	1	0	100	0	1
IGF2	0	152	7	1	0	156	3	1
INS	2	77	0	0	2	77	0	0
MOTI	1	85	0	0	1	85	0	0
NEU2	1	135	2	0	1	136	1	0
NEUT	2	137	2	0	2	137	2	0
NMU	2	125	4	1	3	129	0	0
NPY	1	62	1	0	1	63	0	0
PACA	3	139	5	1	3	142	2	1
PAHO	1	49	1	0	1	49	1	0
PTHR	2	133	12	0	2	142	3	0
PTHY	1	86	3	0	1	88	1	0
SECR	2	126	1	0	1	127	0	1
SMS	1	86	0	1	1	86	0	1
TRH	0	218	14	0	0	220	12	0
TSHB	0	112	2	0	0	113	1	0
VIP	2	162	5	2	2	164	3	2
Total	**36**	**2811**	**75**	**13**	**35**	**2851**	**35**	**14**
Sensitivity	73.5%				71.4%			
Specificity	97.4%				98.8%			
CCR[2]	97.0%				98.3%			

[1]: TP: true positives; TN: true negatives; FP: false positives; FN: false negatives; positives=cleavage sites; negatives=non-cleavage sites

[2]: Correct classification rate

APPENDIX: SUPPLEMENTARY MATERIALS[1]

Table S1a. Evidence supporting five previously unreported neuropeptide genes in the chicken EST database

Table S1b. Evidence supporting five previously unreported neuropeptide genes in the chicken genome

Table S1c. Evidence supporting five previously unreported neuropeptide genes in the chicken high-throughput genome sequence

Table S2. Comprehensive distribution of neuropeptide and convertase gene EST across tissues and stages

Table S3. Comprehensive description of the 22 chicken microarray experiments analyzed

Table S4. Neuropeptide and convertase gene differential expression *P-values* across 22 microarray studies grouped by tissue type

[1]: The supplementary materials can be accessed in Neuropeptides from the manuscript: Delfino, K., Southey, B., Sweedler, J., and Rodriguez-Zas, S. Genome-wide census expression profiling of chicken neuropeptide and prohormone convertase genes). In Press as of November 20, 2009.